PSYCHOMETRIKA—VOL. 11, NO. 3
SEPTEMBER, 1946

AN APPROACH TO THE PROBLEM OF DIFFERENTIAL PREDICTION

HUBERT E. BROGDEN
THE ADJUTANT GENERAL'S OFFICE*

A procedure for maximizing selective efficiency is developed for application to situations in which it is desired to select from a single group of applicants for several possible assignments. The problem of comparable units for the several criteria whose values must be compared to each other for differential assignment purposes is discussed. It is demonstrated that, assuming linear regressions, maximal selection is obtained if individuals in any given assignment are differentiated from those rejected according to critical rejection scores on the multiple weighted sum of the predictors and from another possible assignment by critical difference scores which are merely the differences between the two critical rejection scores. Since the relationships just indicated give no way of determining the magnitude of the critical scores required to select the required number of persons for each assignment, a successive approximation procedure for accomplishing this purpose has been devised and a computational example is worked out.

The procedures for obtaining maximum efficiency in selecting personnel by means of test scores or other predictors are simple and well known when a single assignment is involved. So far as the author is aware, no procedure has been devised for maximizing efficiency of selection and assignment when each individual may be eligible for several assignments. The present paper will be concerned with presentation of such a procedure.

Before attempting to formulate the problem in mathematical terms, the question of comparability of the units for the criteria of the several assignments will be given some consideration. While criteria in standard score form might be regarded as comparable, this solution involves the tacit and undesirable assumption that all criteria are equal in both variability and importance. In certain assignments the nature of the work may be such that all individuals produce very nearly the same amount, while in other assignments considerable variation may occur. It would, of course, be advantageous to place an individual equally good at both types of work in the latter assignment. Similarly, jobs vary in importance to over-all effici-

* The opinions expressed are those of the author and are not to be construed as official or as those of the War Department.

ency of the organization. Thus, if an applicant for employment in a newspaper office were highly but equally skillful at both sweeping floors and operating a linotype machine, it would be desirable to have him operate the linotype machine. Variation in the efficiency with which the linotype machine is operated affects over-all efficiency of the newspaper office much more than variation in the efficiency with which floors are swept. The following discussion of the characteristics of a meaningful criterion is pertinent to the problem of the comparable units and leads rather directly to the solution which seems to the author to be most desirable. While it will be assumed in this discussion of criteria that the classification problem is that of an industrial concern, most of the comments are with some modification pertinent to classification problems of organizations such as the United States Civil Service Commission or the Army.

Although standard scores were considered defective from the viewpoint of comparability, it is definitely desirable to employ mean deviates, and it will be assumed hereafter that all variables are expressed in such terms. The only reasonable alternative to mean deviates would be variables expressed in terms of absolute zero. Apart from the impracticability of attempting to determine absolute zeros in our present stage of development of selection procedures, the information made available by determining absolute zero is not directly pertinent to the problem of making the best possible selection and assignment from the available applicants. In most selection problems the essential comparison is between the given individual score and the expected score if one were to choose at random from the sample of applicants. This expected score is of course the mean. Selection of an individual producing ten units more than the average applicant would effect a saving of exactly ten units no matter whether the mean were twenty above absolute zero or one hundred above it.

Usually in selection problems a test or a battery of tests is employed to identify in a group of applicants those individuals who will perform most efficiently on the job. Presumably if time and expense were not important, the criterion itself would be employed as the selector. Ideally, a criterion would indicate the difference between the cost to the employer per unit of produce or per service rendered by the given individual and the average applicant (that is, it would be expressed in mean deviate form) multiplied by the number of each of the various types of units called for in the job that the given individual could be expected to produce in a given time unit. The criterion should allow for errors, training costs, turnover, or for any additional cost accounting factors which might be related to individual differences in the abilities or traits of the persons on whom the criterion scores are obtained. For example, certain types of overhead would be reduced with more efficient producers in some instances. Thus the criterion would indicate the total saving (or loss) to be obtained by selecting the given instead of the average applicant. If the criterion were expressed directly in terms of a reliable measure of the saving in dollars, the obtained criterion scale would be in units having the same meaning at all points of the scale. In addition, units of several criteria so expressed would have the same meaning, or we might say, selective significance, and could be directly compared not only with other units on the same criterion scale, but with any unit on any criterion scale. That is, if individual X in Job A could be expected to perform such that a saving of one hundred dollars would be expected (over the average individual) and in Job B such that a saving of two hundred dollars would be effected (the criterion values would be +100 and +200, respectively), it could be said that a saving of one hundred dollars would be obtained by placing the individual in Job B. If, furthermore, successive pairs of individuals were to have criterion values of —100 and —98, 105 and 107, and 95 and 97 in Job A; and —22 and —20, 117 and 119, and 205 and 207 in Job B, the differences in desirability for employment between these successive pairs would be exactly the same as far as the employer is concerned. In that sense, then, it can be said that the criterion units are comparable.

It is realized that the reasoning here is from the employer's viewpoint only, and that there are, even so, many intangibles that would never be expressed in monetary terms. In fact, a close approximation to such a criterion could probably not be obtained. However, consideration of the desirable characteristics of the criterion may help in obtaining the best approximation possible under the given circumstances. The desirability of obtaining such an approximation would be much more evident in a differential prediction problem than in those involving a single criterion, since, in the former, the problem of comparing units of different criteria is added to that of comparing units at different parts of the scale of the same criterion. Insofar as the units and scales are not comparable in the way in which we have defined comparability, the selection procedure to be described will not obtain maximal results. Very probably, it would be desirable to employ weights determined by subjective judgment as to the importance of the job rather than to employ ratings or other such criteria in raw or standard score form. However, it should possibly be emphasized that the form of the procedures to be developed here are not dependent upon the character of the units employed, even though the results themselves may be considerably affected.

If we assume that our criteria are expressed in terms of dollars as units, the object of the selection procedure would be to maximize the saving. Procedures for accomplishing this will be developed in the following presentation. Let us assume that:

1. All zero-order and partial regressions are linear.
2. All predicted criterion values have been computed for all criteria for the same battery of tests. The symbol $\bar{y}_i$ will refer to a predicted value of any given criterion i.
3. All statistical constants refer or apply to the sample of applicants.

It may be readily shown—simply by summing the various arrays—that

$$\sum_s \bar{y}_i = b_{yx} \sum_s x \, , \tag{1}$$

where the subscript s indicates that summation is within those above a point of cut *on the predictor.* If the criterion is expressed in terms of dollars saved, it can be seen that $\sum_s \bar{y}_i$ gives the amount saved for that selected group by the selection process.

With n assignments, the total saving in dollars (the criterion) would be

$$I = \sum_s \bar{y}_1 + \sum_s \bar{y}_2 + \cdots + \sum_s \bar{y}_i + \cdots + \sum_s \bar{y}_n \, . \tag{2}$$

I, the total saving in dollars, is the index to be maximized.

Note: The reader is reminded that, if no bias enters into the prediction, the algebraic sum of the errors of prediction approaches zero, so that $\Sigma \bar{y}$ equals Σy. Note though that $\sigma_{\bar{y}}$ equals $r_{xy}\sigma_y$ or, in terms of multiple prediction $\sigma_{\bar{y}}$ equals $R_{y.x_1x_2\ldots x_n}\ \sigma_y$. This latter point is significant. We have already indicated that σ_y, if determined in proper units, would increase as the "importance" of the job increases. Since in practice we must employ test scores as predictors, Σy must be calculated for those selected by the tests. This is equivalent to selecting on $\bar{y}$. Hence, the fact that $\sigma_{\bar{y}}$ is a function both of σ_y and the multiple correlation means that the accuracy with which the criterion can be predicted will have considerable effect upon the differential assignment of the applicants, since use of y instead of y values is equivalent to weighting according to the size of the correlation. That the weighting for purposes of differential prediction is properly a function of R should, in any event, be apparent. If the multiple correlation is zero for a given assignment, the test scores are completely unrelated to the criterion values and to the desirability of selecting individuals obtaining these scores. Hence, it is apparent that the assignment involved should be completely disregarded in selecting men for other assignments even though it may be far more important to obtain men high in that assignment than in the case of any remaining assignments. While the significance of this general principle is most evident in differential prediction, it has definite implications in employing a single predictor to select for several assignments.

The problem is that of defining the bounding surfaces distinguishing between the various assigned groups and differentiating the assigned groups from the unassigned in order to obtain the desired maximized value of I. While an exact general solution would involve further and highly restrictive assumptions (i.e., normality of the correlation surfaces) concerning the nature of the frequency functions and would in any event be exceedingly complex, we shall be content in the present paper with demonstrating that I will be maximized if arbitrarily determined "critical rejection scores" on the $\bar{y}$ values are employed to distinguish between assigned groups and those rejected entirely, while the differences between all possible pairs of such critical rejection scores are employed as "critical difference scores" on the "difference variables" to distinguish between the various possible assignments. It will be noted that the proposed solution requires that three separate propositions be demonstrated:

1. The desired bounding surface differentiating between the assigned group i and the rejected group is defined by a given critical rejection score on $\bar{y}_i$ (note the direct implication that the bounding surface is not curvilinear).
2. The desired bounding surface between any two assigned groups such as i and j is adequately defined by a critical difference score on the difference variable $(\bar{y}_i - \bar{y}_j)$.
3. The exact desired critical difference score on $(\bar{y}_i - \bar{y}_j)$ is the difference between the two critical rejection scores involved.

No equation for directly determining the exact critical rejection scores is to be developed although a method of successive approximations will be suggested.

With reference to all of the propositions to be proved, it will be helpful to note that it is axiomatic that I is maximized when (1) all $\bar{y}$ values for any rejected individual are equal to or lower than any $\bar{y}$ value for assigned individuals on the criterion of their assignments and (2) it is not possible to replace individuals who have higher $\bar{y}$ values on other than the criterion of their assignment with unassigned individuals such that the loss effected by replacement is smaller than the gain effected by change in assignment, or it is not possible to effect a gain by any combination of such replacements and reassignments.

Although the first of our propositions must be true if the assumption of linear regression is met, this relationship between the assumption and the proposition is not immediately evident. Suppose, in a multidimensional space with the $\bar{y}$'s as coordinates, we were to prepare separate plots of individuals having successive $\bar{y}_i$ scores from

—3.0 up to +3.0. That is, the first plot would consist of all individuals having a $\bar{y}_i$ score of —3.0, the second of individuals having a $\bar{y}_i$ score of —2.9, etc. Each of these plots would constitute a "slice" through the hyperspace in which the individuals were plotted. It is quite apparent that our rejection bounding surface, that is, our means of distinguishing between those to be rejected and those belonging in the given assignment, must be one of these "slices," since within each "slice" all $\bar{y}_i$ values are the same and within the "slices" above or below all $\bar{y}_i$ values are, respectively, larger or smaller. In other words, it makes no difference in summing the $\bar{y}_i$ values which of a group of individuals having a constant $\bar{y}_i$ score are chosen for rejection or for assignment elsewhere and, in addition, all those having such a constant score are to be preferred to those having a lower score. Hence, within any given "slice" it makes no difference in maximizing $\sum_s \bar{y}_i$ which individuals are selected for other assignments, although the relative proportions chosen from successive slices do very definitely affect $\sum_s \bar{y}_i$. Among those not assigned elsewhere, a lower value would never be selected in preference to a higher $\bar{y}$ value, no matter what scores are obtained on criteria of other assignments. To state otherwise would be to imply that $\bar{y}_i$ is not the best prediction of y_i. In other words, the fact that we are concerning ourselves with $\bar{y}_i$ instead of y_i, the criterion itself, does not influence the validity of the foregoing statement so long as the $\bar{y}_i$ values represent the "best" prediction obtainable from the test scores in terms of which the individuals are plotted. From our assumption of linear regression lines it may be stated, in any event, that the mean criterion score is equal to the mean predicted criterion score for any segment of our "slice." From this assumption it also follows that none of the "slices" is curvilinear. As these "slices" become infinitesimally thin they are defined exactly by a critical score on $\bar{y}_i$, since they consist of individuals having the same constant $\bar{y}_i$ score. Note that no statement concerning the magnitude of the critical score has been made. It is merely shown that the bounding surface is defined by some critical score on $\bar{y}_i$. Note also that the bounding surface is perpendicular to the multiple regression line determining the given $\bar{y}_i$ values. Thus we have demonstrated the first of our three propositions, namely, that the desired bounding surfaces differentiating between an assigned group i and the rejected group is defined by a given critical score. It would possibly be more appropriate to say that we have shown that it follows directly from our assumption of linearity of regression lines and have elaborated somewhat its meaning and implications.

The demonstration of the second proposition follows that of the first almost exactly. If we were to plot individuals having the same difference scores $(\bar{y}_i - \bar{y}_j)$ we would obtain a series of "slices" or bounding surfaces analogous to those obtained by "slicing" on $\bar{y}_i$. From our assumptions of linearity of regression lines these bounding surfaces would also be linear. The necessity that each bounding surface differentiating between any two assignments must be defined by a constant value of difference variables $(\bar{y}_j - \bar{y}_k)$ is evident when it is realized that the differences between the $\bar{y}$ values are a direct indication of the gain or loss in I to be effected by shifting individuals from one assignment to another. Thus the second of the three points to be proved also follows almost directly from the assumptions of linearity of regression lines.

The next and third point to be demonstrated is that I is maximized when each critical difference score is the difference between the critical rejection scores on the $\bar{y}_i$ values of the two assignments involved, assuming that the proportion of individuals in each assignment remains constant. Suppose we consider the plane formed by plotting paired $\bar{y}_i$ and $\bar{y}_j$ values. To obtain agreement with our first proposition, assigned groups i and j must be separated from those rejected by critical scores on $\bar{y}_i$ and $\bar{y}_j$. We will refer to these critical scores as ${}_c\bar{y}_i$ and ${}_c\bar{y}_j$. The rejection boundaries defined by the critical scores would be straight lines, each perpendicular to its corresponding axis. To obtain agreement with our second proposition the two assigned groups must be separated from each other by one of a family of lines defined by constant values of the difference $(\bar{y}_i - \bar{y}_j)$. Parenthetically the reader is reminded that the slope of all members of this family of lines is the ratio between the s.d.'s of $\bar{y}_i$ and $\bar{y}_j$, while the s.d.'s are in turn equal to the multiple correlation of the several predictors for the given assignment times the s.d. of the criterion of that assignment.

Suppose that the third proposition—which remains to be demonstrated—has been assumed valid in separating the three groups and the boundary between the two assigned groups is that one of the family of difference lines identified by the difference between the two critical rejection scores or ${}_c\bar{y}_i - {}_c\bar{y}_j$. If this were the case, the three boundaries would have a common point of intersection. Since the critical difference score may be increased or decreased only, and since the critical rejection scores are determined after each such change in the critical difference score by the requirement that the number in each category remains constant, it is clear that only these two types of changes are possible. Consider Figure 1. Here, we have assumed that on the plane formed by $\bar{y}_i$ and $\bar{y}_j$ the critical rejection score for $\bar{y}_i$ is 2.0; the critical rejection score is 1.0 for $\bar{y}_j$, and the critical difference score is 1.0. These are indicated by the heavy lines. Suppose that the critical difference score were increased so as to transfer m individuals

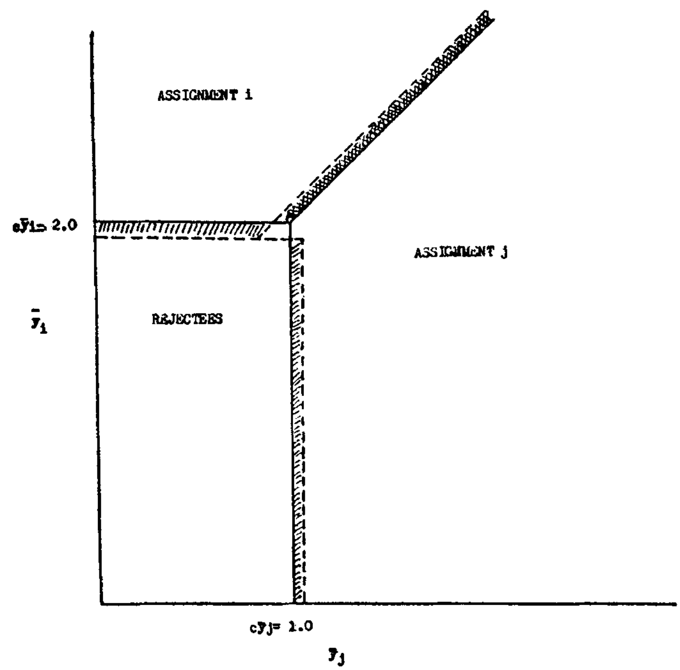

FIGURE 1

from assignment i to assignment j and changes were made in the critical rejection scores so as to allow no change between the number in assignment i and j before the transfer and the number obtained afterwards. The three shaded areas between the pairs of heavy and dotted lines delineate the individuals whose assignment is changed. It is obviously the resulting changes in the criterion of their assignment and the consequent change in the criterion scores of these individuals that will change I. In the shaded area between the two critical difference scores the m individuals involved would have been shifted from assignment i to assignment j. Since, in the case of those individuals directly on the original rejection lines (the heavy lines in Figure 1), the difference between the $\bar{y}_i$ and $\bar{y}_j$ scores was 1.0, the shift of each such individual to assignment i would decrease the over-all sum by exactly 1.0, since the criterion score for such an individual on criterion i, their assignment after the shift, is exactly 1.0 larger than their score on criterion j, or that for their previous assignment. However, the average difference in criterion scores of the individuals shifted is somewhat larger than 1.0, since the shift increased the difference critical score. Hence the decrease effected in each individual's criterion score would be 1.0 plus an increment which we shall call Δd. The effect of this shift on the over-all sum would be to add, algebraically, $-m\ (1.0 + \Delta d)$. The individuals in the shaded area between the heavy and dotted line indicating the two critical rejection scores on $\bar{y}_j$ are being shifted to the rejected groups. Since, on the average, their $\bar{y}_j$ scores are somewhat above 1.0, this shift adds $-m(1.0 + \Delta\bar{y}_j)$ to I. By analogous reasoning, the inclusion of those in the third shaded area adds $m(\ 2.0 - \Delta\bar{y}_i)$ to I. The net change is then $-m(1.0 + \Delta d) - m(1.0 + \Delta\bar{y}_j) + m(2.0 - \Delta\bar{y}_i)$, which reduces to $m(-\Delta d - \Delta\bar{y}_i - \Delta\bar{y}_j)$, since m was made constant in order that the total in each assignment would remain the same.

Figure 2 is to be interpreted in the same manner as Figure 1. The shading indicates the areas affected by the changes in the critical difference scores and compensating changes in the critical rejection scores. In the shaded area between the two critical difference scores the mean difference is somewhat smaller than 1.0, so that in shifting individuals to assignment i, the change effected in I would be $m(1.0 - \Delta d)$. In the same way it can be seen that the shaded area between the heavy and dotted lines for $\bar{y}_i$, which represents individuals who have been shifted from assignment $\bar{y}_i$ to rejection, have an average score of something over 2.0, or say $2.0 + \Delta\bar{y}_i$. The loss occasioned by shifting to the rejected group is $-m(2.0 + \Delta\bar{y}_i)$. In-

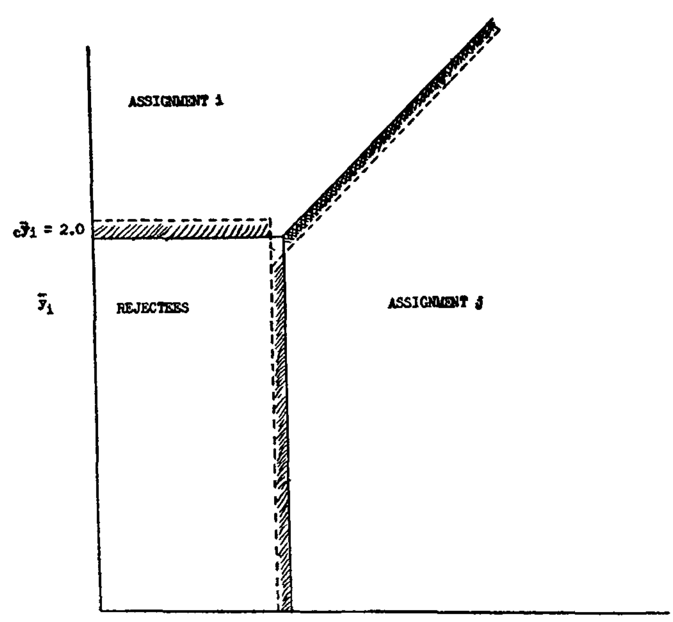

FIGURE 2

dividuals in the corresponding shaded area for $\bar{y}_j$ have an average criterion score of $1.0 - \Delta\bar{y}_j$ so that the effect of changing these individuals from the rejected group to assignment j is to add $m(1.0 - \Delta\bar{y}_j)$ to I. The total of these several changes is again $m(-\Delta d - \Delta y_1 - \Delta y_2)$. The area in which individuals formerly rejected are now given assignment j in Figure 1 and the area in Figure 2 where individuals formerly rejected are given assignment i have been neglected, since they are second-order differentials. It is true, of course, that the changes in the critical scores of $\bar{y}_i$ and $\bar{y}_j$ required in order to hold the proportion of cases in each assignment constant will mean that in other planes the critical difference score for, say, $\bar{y}_i$ and $\bar{y}_k$ or $\bar{y}_j$ and $\bar{y}_k$ is no longer equal to the difference between the critical scores on the two criteria defining the particular dimension in question and, it might be argued, the changes in the critical scores of other variables that will be required in order to hold the number in each assignment constant might raise I rather than lower it. However, such changes are exactly analogous to those we have discussed, and the foregoing proof will apply in the case of these further changes.

Since changes in critical scores from our proposed maximal position can be of two types only, and since it has been shown that with each of these changes a decrease in I is effected, our third proposition has been demonstrated. When all critical difference scores are equal to the differences between the two critical rejection scores, no change in critical difference scores which maintains the same proportion in each assignment will effect an improvement in I.

Demonstration of these three propositions means that with selection of any set of critical rejection scores, determination of critical difference scores from the rejection scores and assignment will be such that I is maximized for the proportions of cases in each of the assignments. The selection of the most feasible procedures of determining critical scores such that the proper number of individuals are allocated to each assignment is the principal remaining problem. It has already been indicated that a theoretical solution to this problem which would have general application does not appear possible, since assumptions concerning the nature of frequency solids formed by plotting individuals on $\bar{y}$-axes would have to correspond rather closely to the empirical data in order that such a solution would be useful. In any event the solution would be exceedingly complex and very probably rather cumbersome from the computational viewpoint. The alternative would appear to be a successive approximation procedure

A definite step-by-step successive approximation procedure will be presented. Although this procedure is not necessarily the most rapid, it is simple and easy to apply. Let us assume that we are starting with a list of $\bar{y}$ values for all assignments and all individuals. The actual steps of the proposed procedure are as follows:

1. Prepare frequency distributions of $\bar{y}$ values for each assignment, recording the identification number of each individual instead of a frequency tally.
2. Draw a line through the top part of the frequency curve isolating, in the segmented portion, the required number of cases for that assignment. The point of cut of this line is the first estimate of the critical rejection score (designated ${}_c\bar{y}_i$).
3. Compute critical difference scores (${}_c\bar{y}_i - {}_c\bar{y}_j$).
4. Identify in each frequency distribution, by referring the identification number of selected individuals to the listing of $\bar{y}$ values and to the critical rejection scores, all individuals above ${}_c\bar{y}$ in other frequency distributions.
5. At the time that the $\bar{y}$ values are located in the listing, refer to the critical difference scores and determine in which assignment the individual belongs. Indicate in some way on the individual's identification number in the frequency distribution that he is or is not to be included in that assignment.
6. Lower the critical scores to include enough additional subjects to replace those assigned elsewhere.
7. Determine whether any of the new assignees were included (either in the first or second selection) in any other assignment.

TABLE 1

Listing of $\bar{y}$ Values in Numerical Example

Man No.	$\bar{y}_i$ Values $\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_4$	Man No.	$\bar{y}_i$ Values $\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_4$
1	—.8	—2.4	.1	—.4	11	—.7	—1.2	.3	.1
2	—.3	—1.9	.3	.7	12	—.9	—1.1	—.7	—1.4
3	—.4	—1.8	.1	.3	13	—.3	—1.1	.1	—.9
4	—1.1	—1.7	.6	1.1	14	—.8	—1.0	—.3	.5
5	—.1	—1.6	—.8	—1.0	15	—.4	—1.0	.6	0.0
6	—.5	—1.5	.7	—.3	16	—.3	—1.0	—.5	.3
7	—.7	—1.4	—.1	—1.3	17	.1	—.9	—1.0	2.4
8	—.6	—1.3	.2	.4	18	—.5	—.9	.8	.9
9	—.1	—1.3	—.6	—.4	19	.2	—.9	.5	—.5
10	—.2	—1.2	—.9	—1.1	20	—.2	—.8	.2	1.2

Table 1 (Continued)

Listing of $\bar{y}$ Values in Numerical Example

Man No.	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_4$	Man No.	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	$\bar{y}_4$
21	— .7	— .8	.6	.9	61	— .5	.3	.4	1.0
22	.6	— .8	— .2	.9	62	— .1	.3	.3	0.0
23	.3	— .7	.2	— .1	63	— .3	.3	— .4	—1.8
24	.7	— .7	— .6	.2	64	.1	.3	.2	— .9
25	— .2	— .7	.3	.6	65	0.0	.4	— .4	— .2
26	.2	— .6	.2	.8	66	.6	.4	.1	.4
27	— .9	— .6	.1	— .3	67	1.0	.4	0.0	1.8
28	— .1	— .6	— .2	— .5	68	.4	.4	0.0	0.0
29	— .6	— .5	—1.1	—1.7	69	.6	.5	— .7	— .8
30	.5	— .5	— .5	— .9	70	.4	.5	1.1	1.3
31	.2	— .5	— .1	.2	71	— .2	.5	.1	— .8
32	— .2	— .5	0.0	1.0	72	.9	.5	0.0	— .6
33	— .3	— .4	.3	— .7	73	— .1	.6	0.0	1.2
34	— .6	— .4	.4	1.5	74	0.0	.6	.9	1.9
35	.1	— .4	.5	.3	75	.1	.6	.4	.4
36	.4	— .4	.7	.1	76	.8	.7	.6	—2.4
37	— .5	— .3	— .1	— .6	77	.1	.7	— .4	— .1
38	— .4	— .3	0.0	.4	78	0.0	.7	.1	.5
39	.7	— .3	— .1	.6	79	.1	.8	.3	1.0
40	.1	— .3	— .3	—1.0	80	0.0	.8	— .5	— .3
41	— .2	— .2	.4	1.7	81	.1	.8	— .1	0.0
42	— .3	— .2	.3	.8	82	.8	.9	0.0	— .1
43	—1.0	— .2	.2	— .2	83	— .2	.9	— .2	.2
44	— .1	— .2	.7	— .7	84	.3	.9	— .6	.5
45	— .1	— .1	— .3	— .2	85	0.0	1.0	— .6	—1.3
46	.2	— .1	— .9	— .6	86	.2	1.0	— .4	.5
47	— .1	— .1	.5	—1.0	87	.5	1.0	— .1	— .2
48	— .4	— .1	1.0	— .5	88	.4	1.1	— .2	—1.2
49	.2	0.0	— .2	— .7	89	.3	1.1	— .3	— .4
50	.3	0.0	.5	1.6	90	0.0	1.2	.8	1.3
51	.6	0.0	— .1	—1.6	91	.9	1.2	0.0	1.1
52	0.0	0.0	.2	— .5	92	0.0	1.3	— .2	—1.5
53	.3	.1	.4	— .4	93	1.1	1.3	— .1	— .1
54	.5	.1	.1	.8	94	.1	1.4	— .3	—1.9
55	.7	.1	0.0	.6	95	— .4	1.5	— .3	.3
56	.3	.1	.4	.7	96	.2	1.6	— .2	— .8
57	— .3	.2	— .3	—1.1	97	0.0	1.7	— .4	.2
58	— .6	.2	— .7	— .3	98	.3	1.8	.1	1.4
59	— .1	.2	— .8	.7	99	0.0	1.9	.1	.1
60	.4	.2	— .5	—1.2	100	.5	2.4	0.0	.1

8. If so, repeat steps 5, 6, and 7 until the required number are in each assignment and no individual is in more than one assignment.

A numerical example will serve to illustrate this procedure. Let us assume that 13, 15, 5, and 15 per cent of the population of applicants are desired for, respectively, assignments 1, 2, 3, and 4. The numerical values of the $\bar{y}$ values are given in Table 1 and the frequency distributions to be prepared as step 1 are given in Table 2. The heavy lines on the distributions indicate the first selection. The critical scores are, respectively, .6, 1.0, .8, and 1.0. This completes step 2. The critical difference scores obtained as step 3 are:

Assignment i	*Assignment j* 2	3	4
1	.4	.2	.0
2		—.2	—.4
3			—.2

Note that the order of the two variables in subtracting determines the sign of the difference. Assignment is made to the first of the two assignments when the value of the difference variable exceeds that of the difference score.

Identification numbers appearing in more than one frequency distribution have been primed. A line has been ruled through numbers when assignment is elsewhere (Table 2). This completes step 5.

Beyond the first approximation and reassignment, little additional labor is involved in at least this numerical example. In assignment 1, man number 67 (the only individual reassigned) was replaced by man number 30. In assignment 2, men numbers 90, 91, and 93 were reassigned and replaced by men numbers 83, 84, and 85. In assignment 3, man number 74 was reassigned and replaced by man number 90. However, after reapproximating the critical difference score between assignments 3 and 4, man number 90 was assigned to 4 (or remained there). Hence man number 6 was the eventual replacement. In assignment 4 men numbers 70, 91, and 98 were reassigned and replaced by men numbers 18, 21, and 79. However, 18 and 21 were assigned to 3 and 1, respectively, and after the critical difference scores were reapproximated they still remained there. Hence man number 26 was placed in assignment 4.

If it is desired to speed the procedure by making subjective allowances for additional factors in arriving at approximations to the critical scores, the following general principles may be helpful:

1. Assignments whose $\bar{y}$ values have relatively large standard

deviations will tend to have high critical scores, since the relative proportion of the individuals in the overlapping areas, i.e., areas above the critical scores of any pair of $\bar{y}$ values, assigned to a given category will depend upon the relative size of the s.d.'s of the $\bar{y}$ values.

2. Since degree of overlapping is a function of degree of intercorrelation, all critical scores will increase as the degree of intercorrelation of $\bar{y}$ values decreases.
3. In making approximations beyond the first, it should be remembered that when shifts are made in critical rejection scores so as to increase the number in a given category, the changes in the critical difference scores which automatically follow will decrease the number in the remaining assignment.

It is quite possible that the approximation procedures would be much more laborious if critical scores were lower and/or the size of the sample larger than in the numerical example. The author does not feel that it is feasible to offer definite information concerning the labor required in the various types of situations in which the procedures might be applied. The problem illustrated was, however, changed to the extent of requiring 25% in each assignment and the solution obtained by the author in less than one hour's time.

TABLE 2

Frequency Distributions of $\bar{y}$ Values for Four Hypothetical Assignments*

$\bar{y}$	Assignment 1										Assignment 2			
2.4											100'			
1.9											99			
1.8											98'			
1.7											97			
1.6											96			
1.5											95			
1.4											94			
1.3											92	~~93'~~		
1.2											~~90'~~	~~91'~~		
1.1	93'										88	89		
1.0	~~67'~~										85	86	87	
.9	72	91'									82	83	84	
.8	76	82									79	80	81	
.7	24	39	55								76	77	78	
.6	22	51	66	69							73	74	75	
.5	30	54	87	100							69	70	71	72
.4	36	60	68	70	88						65	66	67	68
.3	23	50	53	56	84	89	98				61	62	63	64
.2	19	26	31	46	49	96	86				57	58	59	60
.1	17	35	40	64	79	94	75	77	81		53	54	55	56
.0	74	85	90	99	78	52	65	80	92	97	49	50	51	52
- .1	5	9	28	44	45	47	59	62	73		45	46	47	48
- .2	10	20	25	32	41	83	71				41	42	43	44
- .3	2	13	16	33	42	57	63				37	38	39	40
- .4	3	15	48	38	95						33	34	35	36
- .5	6	18	37	61							29	30	31	32
- .6	8	29	34	58							26	27	28	
- .7	7	11	21								23	24	25	
- .8	14	1									20	21	22	
- .9	12	27									17	18	19	
-1.0	43										14	15	16	
-1.1	4										12	13		
-1.2											10	11		
-1.3											8	9		
-1.4											7			
-1.5											6			
-1.6											5			
-1.7											4			
-1.8											3			
-1.9											2			
-2.4											1			

* Identification numbers are listed in the place of frequency tallies.

TABLE 2 (Continued)

Frequency Distributions of $\bar{y}$ Values for Four Hypothetical Assignments*

$\bar{y}$	Assignment 3										Assignment 4			
2.4											17			
1.9											74'			
1.8											67'			
1.7											41			
1.6											50			
1.5											34			
1.4											~~98~~'			
1.3											~~70~~'	90'		
1.2											20	12		
1.1	70'										4	~~91~~'		
1.0	48										32	61	79	
.9	83	74'									18	21	22	
.8	18	90									26	42	54	
.7	6	36	44								2	56	59	
.6	4	15	21	76							25	39	55	
.5	19	35	47	50							14	78	86	84
.4	34	41	53	56	61	75					8	38	66	75
.3	2	11	25	33	42	62	79				3	16	35	95
.2	8	20	23	26	43	52	64				24	31	83	97
.1	1	3	13	27	54	98	71	66	78		11	36	100	99
.0	32	55	73	38	67	68	72	82	91	100	15	62	68	81
- .1	7	31	37	39	51	81	87	93	99		23	77	82	93
- .2	22	28	49	88	96	92	83				43	45	87	65
- .3	14	40	45	57	89	94	95				6	27	58	80
- .4	63	65	77	97	86						1	9	53	89
- .5	16	30	80	60							19	28	48	52
- .6	9	24	84	85							37	46	72	
- .7	12	58	69								33	44	49	
- .8	5	59									69	71	96	
- .9	10	46									13	30	64	
-1.0	17										5	40	47	
-1.1	29										10	57		
-1.2											60	88		
-1.3											7	85		
-1.4											12			
-1.5											92			
-1.6											51			
-1.7											29			
-1.8											63			
-1.9											94			
-2.4											76			

* Identification numbers are listed in the place of frequency tallies.