APPLICATIONS OF FRACTILE GRAPHICAL ANALYSIS TO PSYCHOMETRY: I. ITEM ANALYSIS*

Rhea S. Das
Indian Statistical Institute, Calcutta

Psychometry is concerned with the measurement of psychological attributes, with particular reference to measurement by means of tests. These tests, which may be termed alternatively "mental tests", "objective tests", or "psychometric tests", generally possess the folowing characteristics: they are composed of a large number of questions, called items; the questions or items are answered by selecting one of several alternatives; and performance on the test is indicated by a simple or weighted summation of the number of right answers. Psychometric analysis is directed toward answering the question: how well does a test actually carry out its function of measurement? To seek the answer to this question, statistical methods are introduced to examine the adequacy of the items and the reliability and validity of the test (3, 4).

Item analysis is the broad topic covering the evaluation of item adequacy by statistical methods. Commonly used indices of item adequacy are difficulty, i.e., the proportion of the population successfully answering the item, and discrimination, i.e., the relationship between item success and test score (3). These two indices can be integrated if the items are analyzed by the probit technique (2), the former as the test score associated with 50% success (m), and the latter as the rate at which the item discriminates, i.e., the probit regression coefficient (b 1). The computational model uses the x and y variates of test score and item success respectively (see Table 1).

**This is the first out of the two articles contributed by the author. The second article on 'Reliability' will follow in the next issue.*

Table 1: Psychometric Analysis in Terms of Variates x and y

Sl. No.	analysis	variate x	variate y
(0)	(1)	(2)	(3)
1.	Item analysis	test score	item success
2.	raliability: test-retest	test score on first occasion	test score on second occasion
3.	reliability: parallel forms	test score on first form	test score on second form
4.	reliability: split half	score on even-numbered items	score on odd-numbered items
5.	validity	test score	criterion score
6.	age group norms	test score	age

Reliability of a test is often defined as the consistency or dependability of its measurements. The more the measurements are free from chance error, the greater the reliability of the test. A wide variety of empirical formulations exist for estimation of test reliability. A feature common to most of these formulations is that they depend upon the relationship between two measures, say x and y, obtained by administering the "same" test to the same population. These measures, x and y, may be obtained by actually administering the test twice; by administering two different forms of the test; or by obtaining two scores from the test, with half of the items contributing to one score, x, and the other half to the other score, y. The reliability estimates are termed, respectively, test-retest, parallel forms, and split-half reliability. Empirically, the correlation between x and y is then taken as the basis for computing the reliability estimate, or as the estimate itself.

Tests are often used to obtain information which will predict or quickly approximate the true or actual psychological attribute under consideration. Hence, it becomes important to determine the accuracy of the predictions or approximations. The degree of accuracy provides the index of the validity of the test. For this purpose, a more "true" measure of the attribute is chosen as the criterion or standard against which to judge the test, and the relationship between the test, x, and the criterion, y, provides the empirical assessment of test validity, usually as the correlation between x and y.

From this overview of item analysis, reliability and validity, the following observation emerges: in each case, the relationship between two variates x and y is of primary importance. A statistical model which can generally be applied to psychometric analysis of items, reliability, and validity, is provided by fractile graphical analysis.

The method of fractile graphical analysis has been developed and explained by P. C. Mahalanobis (5). Therefore here it may be only briefly reviewed. Two random samples are drawn from a bivariate population of two random variates, x and y. Each unit or member of the population consists of a pair of values for the variates, x and y. Each sample, consisting of n such units, can be arranged in order of increasing value of x, and the ordered units may then be divided into groups of equal number, termed "fractile groups". For each fractile group, the mean value of y is computed. A fractile graph may then be drawn by connecting the mean y values of the fractile groups, which have been placed equidistantly on the x axis. This fractile graph may be drawn for the first sample, for the second sample, and after pooling the two samples, for the combined sample. The "error" associated with the fractile graph is defined as the area contained between the fractile graphs of the two samples. As the two samples can be pooled to form the combined sample, they will be referred to subsequently as subsamples (ss), hence the first sample will be ss_1, and the second sample, ss_2.

For each of the different topics of psychometric analysis, a basic consideration is the degree of error involved in item or test measurement. In fact, as the error is decreased, it may be said that the adequacy or accuracy of the test is increased. The amenability of fractile graphical analysis for ascertaining the error associated with psychometric measurements is suggested by Table 1, in which it is seen that, x and y variates are characteristic of each of the cases listed. In addition to item analysis, reliability and validity, age group norms have also been included in Table 1, as a measure of error would be desirable in evaluating the test performance of different age groups as well as the performance of individuals relative to their age.

Fractile Graphical Analysis of Items:

Item analysis is carried out for a number of purposes, among which are the following:

i. to improve a test for immediate use on the initial group of persons;

ii. to select the best items of a test for the final form after an experimental try-out.

iii. to provide the test constructor with a statistical check on his subjective judgment of the characteristics of the test items;

iv. to set up parallel forms of a test; and

v. to compare the answer patterns of two or more groups for diagnostic or discriminative purposes.

As Guilford has pointed out, a normal ogive relationship generally holds between proportion of item success and test score (3). Treating test score as x, and transforming proportion of item success, y, to probits, this relationship becomes linear (2). Fitting a straight or probit line to the observed data in turn permits estimation of item difficulty as the test score associated with 50% item success, m, and also an index of item discrimination, the regression of the probit line on test score b. Purposes (i) to (v) mentioned in the previous paragraph can be satisfied using this technique for item analysis.

Table 2: Number and Proportion of Right Answers to a Single Item for Population A

group no.	centile	total number of subjects			number right			proportion right		
		ss_1	ss_2	com	ss_1	ss_2	com	ss_1	ss_2	com
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	10	10	11	21	0	0	0	.0000	.0000	.0000
2	20	12	12	24	0	0	0	.0000	.0000	.0000
3	30	12	12	24	0	0	1	.0000	.0000	.0417
4	40	12	12	24	3	2	4	.2500	.1667	.1667
5	50	12	12	24	4	1	6	.3333	.0833	.2500
6	60	12	12	24	4	5	8	.3333	.4167	.3333
7	70	12	12	24	3	5	8	.2500	.4167	.3333
8	80	12	12	24	8	6	14	.6667	.5000	.5833
9	90	12	12	24	7	9	16	.5833	.7500	.6667
10	100	12	12	24	9	9	18	.7500	.7500	.7500
	Total ...	118	119	237						

Items within a test may also be compared in terms of the error associated with the trend as indicated by the combined sample. The error may be conveniently found out using fractile graphical analysis. Here, the basic data are test score, x and item success, y. The fractile groups are formed for test score, x, in the usual way. As the y value for each member of the population is dichotomous, the number of members passing the item, or that number converted to percentage of the group passing the item, may be taken as the value of y associated with each fractile group. The fractile groups and their associated percentages of item success are formed for ss_1, ss_2, and the combined sample. The fractile graphs are drawn, and the error area is measured. It may be noted that percentages or proportions have been used for this purpose, rather than probits. This has been done as 100% item success is often achieved by high scoring persons, and 0% is often associated with low scores. For the probit transformation, these values are infinite, hence, the values are more amenable to graphic treatment in their original form.

Table 3: Number and Proportion of Right Answers to a Single Item for Population B

group no.	centile	total number of subjects			number right			proportion right		
		ss_1	ss_2	com	ss_1	ss_2	com	ss_1	ss_2	com
0	1	2	3	4	5	6	7	8	9	10
1	10	7	8	15	0	0	0	.0000	.0000	.0000
2	20	8	8	16	0	0	0	.0000	.0000	.0000
3	30	8	8	16	0	1	0	.0000	.1250	.0000
4	40	8	8	16	0	1	2	.0000	.1250	.1250
5	50	8	8	16	1	0	1	.1250	.0000	.0625
6	60	8	8	16	1	1	2	.1250	.1250	.1250
7	70	8	8	16	4	1	6	.5000	.1250	.3750
8	80	8	8	16	4	6	9	.5000	.7500	.5625
9	90	8	8	16	6	6	12	.7500	.7500	.7500
10	100	10	9	19	9	9	18	.9000	1.0000	.9474
	Total	81	81	162						

To illustrate the application of the fractile graphical method to item analysis, 16 items from a nonverbal reasoning test of the analogies type, which has been administered to two populations, A and B, were chosen. The analysis is illustrated for one of the items. Table 2 gives the fractile groups, the numbers of subjects in each group for subsamples one and two and the combined sample, the number of subjects answering the item correctly, and the proportion right for population A. Table 3 gives the same data for population B. To illustrate the graphical analysis, Figure I has been prepared for the data given in Table 2. The area contained within the graphs for subsamples one and two was measured using a planimeter. The resulting measurements, in square centimeters, gave the error area, e. The error areas for the 16 items mentioned above are given in Table 4.

Table 4: Average Difficulty (m), Discrimination (b), and Error Area (e) of 16 items for Populations A and B.

Item	population A			population B		
no.	m	b	e	m	b	e
(0)	(1)	(2)	(3)	(4)	(5)	(6)
1	15.000	0.125	10.5	11.500	0.167	5.6
2	26.000	0.114	6.9	25.000	0.111	7.9
3	9.000	0.118	7.4	1.500	0.054	10.1
4	8.500	0.111	6.5	18.000	0.100	7.4
5	12.000	0.103	10.4	16.000	0.095	8.6
6	22.000	0.083	8.4	26.000	0.059	8.4
7	10.500	0.133	4.7	11.500	0.154	8.3
8	16.000	0.182	11.0	16.000	0.148	2.5
9	14.500	0.098	9.1	12.000	0.083	10.9
10	13.500	0.103	16.1	13.500	0.091	4.9
11	15.500	0.133	7.1	14.500	0.143	11.1
12	19.000	0.118	10.5	25.000	0.095	14.6
13	13.000	0.133	6.4	14.500	0.121	15.8
14	17.500	0.100	10.0	10.500	0.133	4.8
15	21.500	0.121	13.3	22.500	0.087	13.5
16	23.500	0.095	14.6	25.500	0.083	7.7

It would be desirable for the estimate of error to be independent of other parameters of the relationship between item success and test score. For purposes of comparison, the difficulty index, m, and discrimination index, b, are presented for each of the items in Table 4 along with the error measurement, e. The data suggest that, for both populations, error as estimated by fractile graphical analysis is unrelated to the item-test score parameters of difficulty and discrimination.

In order to select items within a test for any of the purposes listed previously, the degree of error associated with the item may be considered. If populations A and B are treated as cross-validation samples, the difference between the two values of e as well as their magnitude would serve as the basis for item selection.

References

1. DAS, RHEA S. An alternative approach to item analysis using probit and fractile graphical methods. (Unpublished manuscript).
2. FINNEY, D.J. The application of probit analysis to the results of mental tests. *Psychometrika*, 1944, **9**, 31-39.
3. GUILFORD, J.P: *Psychometric methods*. 2nd Ed. New York: McGraw-Hill, 1954.
4. GULLIKSEN, H. *Theory of mental tests*. New York: John Wiley, 1950.
5. MAHALANOBIS, P.C. A method of fractile graphical analysis, with some surmises of results. *Trans. Bose Research Institute*, 1958, **22**, 223–230.

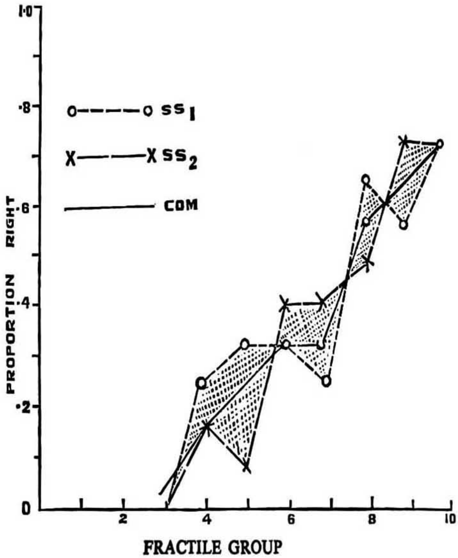

Figure 1. Fractile graph for a single item, giving proportion right of fractile groups arranged in order of test score.