5

ORDER ANALYSIS

Guttman Scaling

Guttman (1944, 1950) described a unidimensional scale as one in which the respondents' responses to the objects would place individuals in perfect order. Ideally, persons who answer several questions favorably all have higher skills than persons who answer the same questions unfavorably. Arithmetic questions make good examples of this type of scale. Suppose elementary school children are given the following addition problems:

(1)		(2)		(3)		(4)		(5)	
	2		12		28		86		228
	+3		+15		+24		+88		+894

It is probable that if subject **A** responds correctly to item 5, he or she would also respond correctly to items 1, 2, 3, and 4. If subject **B** can answer item 2 and not item 3, it is probable that he or she can answer item 1 correctly but would be unable to answer item 4 and 5. By scoring 1 for each correct answer and 0 otherwise, a profile of responses can be obtained. If the arithmetic questions form a perfect scale, then the sum of the correct responses to the five items can be used to reveal a person's scale type. Their sum, is reflected in a series of ones and zeros. In our example:

	Items 1 2 3 4 5	Sum
Subject A has scale type	1 1 1 1 1	= 5
Subject B has scale type	1 1 0 0 0	= 2

Given a perfect scale, the single summed score reveals the scale type. Thus a single digit can be used to represent all the responses of a person to a set of items. Guttman Scales have been labeled deterministic. With five questions and scoring the item as correct or incorrect there are only six possible scale types. These are:

1	2	3	4	5	6
1 1 1 1 1	1 1 1 1 0	1 1 1 0 0	1 1 0 0 0	1 0 0 0 0	0 0 0 0 0

Although there exist 32 possible arrangements (2^5) of five ones and zeros, only six of these form scale types. In general, the **number of scale types for dichotomously** scored data is (K + 1), where K is the number of objects. Although a perfect Guttman scale is unlikely to be found in practice, approximations to it can be obtained by a judicious choice of items and careful analysis of a set of responses to a larger number of items than are to be used in the final scale.

Goodenough's Error Counting

Goodenough revised error counting in Guttman Scaling in a method sometimes known as Scalogram Analysis (Edwards, 1957). First, a set of psychological objects is selected. The objects should be ones that will differentiate participants with varying attitudes or perceptions about the objects along some **single** dimension. Suppose the following six statements have been chosen and 12 student's responses have been obtained in the form of Agreement or Disagreement with these statements. Do these statements constitute a Guttman Scale along the dimension of attitudes toward school? The students are scored (1) for agree and (0) for disagree with the statements. Results are presented in Table 5.1.

Statements	Agree	Disagree
A. School is OK.	______	: ______
B. I come to school regularly.	______	: ______
C. I think school is important.	______	: ______
D. It is nice to be in school.	______	: ______
E. I think school is fun.	______	: ______
F. I think school is better than a circus.	______	: ______

Table 5.1

Item Response Data in the Form of Ones and Zeros

Students	A	B	C	D	E	F	Scores
1	0	1	1	1	1	0	4
2	1	1	1	0	0	0	3
3	1	0	0	0	0	1	2
4	1	1	0	0	0	0	2
5	0	0	1	1	1	0	3
6	0	1	0	1	1	0	3
7	0	1	0	0	1	0	2
8	0	1	1	0	0	0	2
9	1	1	0	1	1	1	5
10	0	1	1	1	0	0	3
11	0	0	1	0	0	0	1
12	0	1	1	0	1	0	3
Sum	4	9	7	5	6	2	33

The basic data are arranged in a table of ones and zeros in which **one (1)** stands for agree and **zero (0)** for disagree, and the rows and columns are summed.

It is initially convenient to rearrange the first table in order of the **row and column sums (see Table 5.2).** Errors in scale types are calculated by subtracting the profile of ones and zeros for each respondent from the perfect scale type with **the same summed** score. For example:

Perfect Scale Type for a score of 5	= 1 1 1 1 1 0
Subject 9 has a score of 5	= 1 1 0 1 1 1
Difference	= 1 −1

The sum of the **absolute value** of each difference is the error. In this case 1 + |−1| = 2 errors.

Table 5.2

Rearrangement of Ones and Zeros

Students	B	C	E	D	A	F	Scores	Errors
9	1	0	1	1	1	1	5	2
1	1	1	1	1	0	0	4	0
2	1	1	0	0	1	0	3	2
5	0	1	1	1	0	0	3	2
6	1	0	1	1	0	0	3	2
10	1	1	0	1	0	0	3	2
12	1	1	1	0	0	0	3	0
3	0	0	0	0	1	1	2	4
4	1	0	0	0	1	0	2	2
7	1	0	1	0	0	0	2	2
8	1	1	0	0	0	0	2	0
11	0	1	0	0	0	0	1	2
Sum	9	7	6	5	4	2	33	20

The total **possible** number of errors is equal to the product of N subjects and K objects (items) or in this case (N)(K) = (12)(6) or 72 possible errors. This is because the maximum possible number of errors is six for any one subject. An estimate of how accurately the particular arrangement approximates a perfect scale is to take the ratio of the found errors to the maximum number of possible errors. Subtracting this ratio from 1.0 renders a coefficient of the scale's ability to reproduce the scores based on the row sums. The **Coefficient of Reproducibility** is :

CR = 1 – (20 Total Errors/72 Possible Errors) = .723.

A **Coefficient of Scalability, CS** = 1 – E / X is determined by finding the ratio of the number of errors (E) to the number of errors by chance (X), subtracted from 1. That is, **CS** = 1 – No. of Errors /.5(N)(K) or 1 – 20 / .5(12)(6) = 0.45. The original items form a poor approximation to a true Guttman scale. **0.60** has been suggested as the lower limit for **CS**. Item C (I think school is important.) accumulates six errors, out of a possible 12, so it is reasonable to eliminate this item. After eliminating item C the responses to the remaining items are reorganized. In this second reorganization (Table 5.3), 12 errors occur.

Table 5.3

Reduced Matrix of One and Zeros

Judges	B	E	D	A	F	Score	Error
9	1	1	1	1	1	5	0
1	1	1	1	0	0	3	0
6	1	1	1	0	0	3	0
12	1	1	0	0	0	2	0
7	1	1	0	0	0	2	0
2	1	0	0	1	0	2	2
4	1	0	0	1	0	2	2
10	1	0	1	0	0	2	2
5	0	1	1	0	0	2	2
3	**0**	**0**	**0**	**1**	**1**	**2**	**4**
8	1	0	0	0	0	1	0
11	0	0	0	0	0	0	0
Sum	9	6	5	4	2	26	12
p	.75	.50	.42	.33	.17		
q	.25	.50	.58	.67	.83		

Note that four errors occur for Judge 3. This person's data should be checked for accuracy in following instructions or for errors in coding. In this new arrangement, a coefficient of the ability of the column sums to reproduce all the responses accurately is **CR** = 1– Errors/(N)(K) or 1 – 12/60) = .80. The **CR** index has been slightly improved by deleting item C from the analysis. **CS** = 1 – 12 / .5(12)(5) = .60. Because further deletion appears not to be of benefit (that is, does not increase the coefficient of reproducibility) no more rearrangements are performed.

A test of the effectiveness of a reproducibility coefficient can be made in light of the minimal reproducibility possible given the average proportion of agree (p) and disagree (q = 1 – p) responses in each column of Table 5.3. By averaging the maximum of p or q in each column, the **Minimal Marginal Reproducibility** (**MMR**) can be obtained. Thus the sum of (.75 + .50 + .58 + .67 + .83) / 5 = .67. The difference between a **CR** of .80 and the **MMR** of .67 is .13. This number is the **Percentage of Improvement (PI).**

It is possible to assign a coefficient of reproducibility to a given respondent. This coefficient may be obtained by subtracting the individuals's profile from a perfect scale vector with the same score. The sum of the absolute differences divided by K items and subtracted from 1 gives the **CR** for the subject chosen. For example if there are nine objects and a single respondents total score is six then:

Objects, K = 9		1 2 3 4 5 6 7 8 9	
Scale type	Y =	1 1 1 1 1 1 0 0 0	Score = 6
Subject's vector	X =	1 1 1 0 1 1 1 0 0	Score = 6
		1 -1	

$\mathbf{CR}_i = 1 - [(1 + |-1|)/9] = 1 - 2/9 = .7778$

Application 1: Cloze Tests in Reading

F. J. King (1974) has utilized scalogram analysis (Guttman Scaling) to grade the difficulty of cloze tests. A cloze test asks the subject to complete passages in which a specific scattered set of words has been deleted. King has suggested that a system for getting children to read materials at their appropriate reading level must be capable of locating a student on a reading level continuum so that he or she can read materials at or below that level. This is what a cloze test is designed to measure.

King constructed eight cloze passages ordered in predicted reading difficulty. If a student supplied seven of 12 words on any one passage correctly, he was given a score of one. If he had fewer than seven correct answers, he received a score of zero. A child's scale score could vary from **zero** to **eight,** a score of 1 for each of the eight passages. If the test passages form a cumulative scale, then **scale scores should determine** a description of performance. A scale score of 4, for example, would have the vector

11110000

and this would indicate that a student could read text material at the fourth level and below. By constructing a table to show the percentage of students at each scale level who passed each test passage, King was able to indicate that "smoothed" scale scores were capable of producing the score vectors with considerable accuracy.

Once the scalability of the passages was determined, King related the reading difficulty of the test passages to the reading difficulty level of the educational materials in general. Thus he was able to indicate which reading material a child could comprehend under instruction.

Application 2: Arithmetic Achievement

Smith (1971) applied Guttman Scaling to the construction and validation of arithmetic achievement tests. He first broke simple addition into a list of 20 tasks and ordered these tasks according to hypothesized difficulty. Experimental forms containing four items for each task were constructed and administered to elementary school children in grades 2-6. Smith reduced his 20 tasks to nine by testing the significance of the difference between the proportion passing for each pair of items. He chose items that were different in difficulty with alpha < .05 for his test. The nine tasks were scored by giving a one (1) each time three or more of the four parallel items for that task were answered correctly and a zero (0) otherwise. These items were analyzed using the Goodenough Technique. The total scale score for each subject was defined as the number of the item that preceded two successive failures (zeros). Thus a subject with the following vector:

Tasks	1 2 3 4 5 6 7 8 9
Subject's vector	1 1 1 0 1 1 0 0 0

would have a scale score of 6.

Table 5.4 shows the coefficients of reproducibility obtained by Smith on the addition tests for two schools and grades 2-6.

Table 5.4

Obtained Coefficients of Reproducibility

Grade	Shadeville	Sopchoppy	Both
2	.9546	.9415	.9496
3	.9603	.9573	.9589
4	.9333	.9213	.9267
5	.9444	.9557	.9497
6	.9606	.9402	.9513
All	.9514	.9430	.9487

The high coefficients of reproducibility indicate that a student's scale score accurately depicts his or her position with regard to the tasks necessary in solving addition problems. For this reason Smiths' results can be used: (1) to indicate the level of proficiency for a given student; (2) as a diagnostic tool; and (3) to indicate the logical order of instruction.

Significance of a Guttman Scale

Guttman has stated that a scale with a **CR** < .90 cannot be considered an effective approximation to a perfect scale. Further study comparing significance tests from Rank Scaling suggests that a **CR** of .93 approximates the .05 level of significance. Other sources suggest that the **CS** should be greater than .60.