Chapter 5
The Guttman Structure and Analysis of Responses

In this chapter, we elaborate the cumulative mechanism introduced in Chap. 2. When items are placed into a single test or questionnaire, and they are considered instances or *manifestations of the same construct*, for example, items of a test of reading comprehension or of a questionnaire on depression, then the responses to the items are generally intended to be summarized by a single score. We have seen that this is assumed in CTT.

The only characteristic that follows from the equation and the conditions in CTT, which assumes that the total score is a sound summary of the responses, is that the theoretical correlations among items are positive and the same, and likewise, therefore, that the theoretical item–total correlations are positive and the same. We may note that, although these correlations in theory are the same, because of sampling variation of one kind or another, in real data they will be different. The sampling variation in CTT, as we have seen, is said to produce error. When correlations are expected to be the same with no sampling error, then they would be expected to be similar, though not identical, in real data which inevitably has sampling error. In this case, it is said that the correlations are assumed or expected to be *homogeneous*. In summary, for a total score to be used to summarize the responses on items of an instrument, it is necessary that the inter-item correlations, and therefore the item–total correlations, are positive and homogeneous.

However, we have also noted that although item facility, or its complement *difficulty*, is considered and described in the CTT context of analyzing data sets, item difficulty is not formalized in any equation of CTT. Indeed, from the perspective of CTT, it is perfectly reasonable to have all items more or less of the same relative difficulty.

In the 1950s, the sociologist and statistician Louis Guttman contributed a great deal to the understanding of the structure of tests. In particular, he introduced the relevance of relative item difficulty in the specification of responses to an instrument and by implication, the operational definition of a continuum. Much earlier in the 1920s and from a different perspective, the psychologist and engineer Louis Thurstone introduced the relevance of item difficulty in the operational definition

D. Andrich and I. Marais, *A Course in Rasch Measurement Theory*,
Springer Texts in Education, https://doi.org/10.1007/978-981-13-7496-8_5

Table 5.1 Guttman pattern

Person response patterns	1	2	3	4	5	6	Total score across items
1	0	0	0	0	0	0	0
2	1	0	0	0	0	0	1
3	1	1	0	0	0	0	2
4	1	1	1	0	0	0	3
5	1	1	1	1	0	0	4
6	1	1	1	1	1	0	5
7	1	1	1	1	1	1	6
Total score across persons	6	5	4	3	2	1	

of a continuum. We will see that this perspective is consistent with that of Rasch measurement theory (RMT).

We now study the structure of responses that Guttman introduced for two reasons; first, because it shows an elaboration of CTT in which relative item difficulty is explicit, second, because it leads into RMT.

Guttman (1950) enunciated the following key requirement that a set of responses should meet *before* the scores on the items could be summed to give a meaningful single score for a person. With respect to dichotomously scored items, and again using proficiency as an example,

> If person A has a greater total score than person B on a test, then person A should have answered all the items correctly that person B has answered correctly, and in addition, some other items that are *more difficult*.

If this condition is held for every pair of persons, then it would provide the perfect *Guttman structure*.

The Guttman Structure

The perfect pattern of the Guttman structure for six items scored 0 for incorrect and 1 for correct is shown in Table 5.1, where the items are ordered in difficulty. In turn, the difficulty of the items can be inferred from the number correct. Clearly, for items, the smaller the total score, the more difficult the item, and for persons, the greater the total score, the more proficient the person. We will see, in the section below, the significance of the ordering of the items on a continuum.

Furthermore, for any total score of a person, the pattern of correct and incorrect responses across the items can be inferred, and all persons with the same total score will have the same pattern. For example, a person with a score of 3 in Table 5.1 will have answered the three easiest items correctly and the three most difficult items incorrectly.

Of course, in real data, two people with the same total score may have different patterns of correct and incorrect responses. CTT does not require the strict ordering of the Guttman structure. However, in the case that items do have different difficulties, the positive correlations among items imply more or less that persons with the same total score will tend to obtain similar scores on the same items. Therefore, if very many people with the same score did have entirely different patterns of responses, then this would be evidence that the scores on the items cannot be summarized meaningfully by a single score. For example, in a test of proficiency with items of different difficulty in which two persons with the same total score have very different patterns, it would imply that one person has answered some more difficult items correctly compared to the other person, and vice versa. As a result, it seems that it is not justifiable to conclude that these two persons have the same proficiency.

The implication of positive item-total correlations between the scores on an item and the total score in CTT are similar though not as strict as in the Guttman structure. In addition, as we have seen, CTT does not rely on relative difficulties of items in its interpretations of a continuum. Therefore, though compatible, there is a major distinction between the implication and interpretations of a Guttman structure and the requirements of CTT.

Interpretations of the Continuum in the Guttman Structure

In general, the Guttman structure provides a tangible interpretation of a continuum, and an understanding of what *more or less* on the continuum means.

The Guttman Structure and Assessment of Proficiency

In a test of proficiency, the relative difficulties of the items provide tangible evidence of relative proficiency implied by a total score. For example, answers to the five questions in Table 5.2 would give that kind of structure.

In the example of Table 5.2, it is very likely to obtain only Guttman patterns. Can you tell why? In addition, from the ordering of the items the proficiency along the continuum is relatively clear. More items could readily be constructed whose difficulties are between those of the difficulties of items in Table 5.2.

In a set of responses to more typical items in a test of mathematics (or other tests), it is unlikely that the perfect Guttman pattern will be found. However, if the items are ordered according to their relative difficulty inferred from their total scores, and the persons are ordered according to their relative proficiency again inferred from their total scores, then it can be expected that there will be predominantly 0s in the upper triangle of the table, and predominantly 1s in the lower triangle. In using the Guttman structure as a framework, it is important to understand some conditions that work against obtaining perfect Guttman patterns in real data. There are two main reasons a Guttman pattern may not be evident.

Table 5.2 An example of a test in mathematics which would result in Guttman patterns

1.	What number does 4 add 3 equal?	$4 + 3 =$ ________
2.	If $x + 5 = 10$, what is the value of x?	$x =$ ________
3.	If $x^2 - 2x + 1 = 0$, what is the value of x?	$x =$ ________
4.	If $a^{x^2-1} = 1$, what is the value of x?	$x =$ ________
5.	If $y = 2x^2$, what is the rate of change of y with respect to x?	$\frac{dy}{dx} =$ ________

(i) One reason is that the items may not be assessing the same variable as expected, which implies that the scores on the items should not be summed to give a meaningful total score. For example, this might happen if items of scientific understanding are mixed up with items of mathematics understanding, understandings that are independent of each other. The analysis is focused on this possibility. Thus, the structure that is expected, the Guttman structure, is a hypothesis for the patterns of observed responses.

(ii) A second reason is that, even if the items do assess the same variable, the items may be very close in difficulty and the persons may all be close in proficiency. For example, within classes in elementary and high schools and other teaching situations, items are often constructed to be of relatively similar difficulty and the people to be tested are usually taught a specific set of material and effectively prepared for the test, and as a result, they have similar proficiencies. In such a case, responses are not affected by relative difficulties of items and relative proficiencies of persons and so any differences in difficulties or proficiencies are essentially effects of random error. Thus not having a Guttman structure is not in itself always evidence that responses should not be summed.

However, ideally in a context where an instrument is being constructed and validated, it is necessary to construct items of different difficulties and to administer them to persons of known differences in proficiencies. Then, if an instrument is given to a homogeneous group of persons in proficiency, and a subset of items is chosen because they are the most relevant, then it would not be necessary to assess the responses according to the Guttman structure. Essentially, here we have distinguished between collecting responses to help construct and evaluate an instrument, and applying a validated instrument to assess persons. We revisit the distinction between the stages of constructing an instrument and using it in assessment again when we study RMT.

Despite the possibility of not having a Guttman structure even when scores on the items of a test can be summed meaningfully, the expectation of the Guttman structure can provide an important framework for both constructing and interpreting the empirical evidence provided by tests. First, if a test constructor, including a teacher of a single class, knows in advance that he or she will use the Guttman structure as a framework for analyzing and interpreting the responses, he or she is likely to write at least some relatively easy and some relatively difficult items. Writing some relatively easy and difficult items, and some that are in between these extremes, helps clarify the substantive proficiencies that reflect more or less of the variable to be assessed. Second, to ensure the best engagement between the students and the items, the items can be ordered in the test from easy to more difficult. It should be clear that an ordering in which difficult items appear early in the test is likely to frustrate students and therefore they are likely to not do as well as they could do on easier items that appear later in the test. We noted in Chap. 2 that the cumulative mechanism of the Guttman structure can be used in attitude measurement.

The Guttman structure is not relevant only for the assessment of proficiency. It can be used in the assessment of attitude. A famous example in the Bogardus social distance scale (Bogardus, 1933) in which attitudes towards different ethnic groups were assessed by questions that reflected different degrees of acceptance and where, in principle, responses would be expected to conform to the Guttman structure.

Elementary Analysis According to the Guttman Structure in the Case of a Proficiency Example

We now consider an analysis of the data in Table 3.1 of Chap. 3 from the perspective of a Guttman structure. Table 3.1 showed responses from 50 persons to a 10-item test. Table 5.3 shows the same data, but the polytomous questions 6, 9 and 10 are now broken down into dichotomous sub-questions. In Table 5.3, the persons are *ordered* according to their *total score*, which is an index of their relative proficiencies and is denoted in the table by R. The items are also *ordered* by their *total score*, which is an index of their relative difficulties.

The responses accord relatively well with the Guttman structure, though not perfectly. Thus as expected, there are mostly 0s in the upper right triangle of this table, and mostly 1s in the lower left triangle. There are also some *anomalies*, for example, person 30 answered item 1, which was the second easiest, incorrectly, and it was the only item the person answered incorrectly. This might be interpreted as a slip by the person, but if possible, should be checked in interview with the student.

In this example, there are no missing responses. The specific place to look for missing responses is for items at the end of the test, which would suggest the students did not have enough time to complete the test. If there were missing responses in the first part of the test, it would also be cause for concern, and would suggest ordering of items in the test was not according to relative difficulty.

Table 5.3 Guttman framework with dichotomous items

Person	Items in order of difficulty																		
	2	1	5	6.4	3	4	6.3	9.1	9.3	6.1	9.2	6.2	7	10.1	8	10.3	10.2	10.4	R	
38	1	0	1	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	6	Lower group (13)
2	1	0	1	1	1	0	1	1	0	1	0	0	0	0	0	0	0	0	7	
40	0	1	0	1	1	1	1	1	0	0	0	1	0	0	1	0	0	0	8	
41	1	1	1	1	1	0	1	1	1	1	0	0	1	0	0	0	0	0	10	
42	1	1	0	0	1	1	1	1	1	1	1	0	0	1	0	0	0	0	10	
35	1	1	1	1	1	1	0	1	1	1	0	1	1	0	0	0	0	0	11	
44	1	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0	0	0	11	
8	1	1	1	1	0	1	1	1	0	1	0	1	1	1	0	0	0	0	11	
9	1	1	0	1	1	1	1	1	1	0	1	0	1	1	1	0	0	0	12	
25	1	1	1	1	0	1	1	0	1	1	1	1	1	1	0	0	0	0	12	
11	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	0	0	0	12	
46	1	1	1	0	1	1	0	1	1	0	1	0	1	1	1	0	1	0	12	
29	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	0	0	0	12	
48	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	0	0	0	13	Middle group (20)
18	1	1	0	1	1	1	1	1	0	1	1	1	1	0	1	0	0	1	13	
32	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	0	0	0	13	
20	1	1	1	1	0	1	0	1	1	1	1	1	1	0	1	1	0	0	13	
13	1	1	1	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	13	
22	1	1	1	1	1	1	1	0	1	1	0	1	1	0	1	0	0	1	13	
34	1	1	1	1	1	1	0	1	1	1	0	1	0	1	1	1	0	0	13	
43	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	13	

(continued)

Table 5.3 (continued)

Person	Items in order of difficulty																		
	2	1	5	6.4	3	4	6.3	9.1	9.3	6.1	9.2	6.2	7	10.1	8	10.3	10.2	10.4	R	
36	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	0	0	0	13	
37	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	0	0	0	13	
27	0	1	1	1	1	0	1	1	1	1	0	1	1	0	0	1	1	1	13	
12	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	1	14	
21	1	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	1	0	14	
14	1	1	1	1	0	1	1	0	1	1	1	0	1	1	1	1	1	0	14	
15	1	1	1	1	1	1	1	0	0	1	1	1	0	1	1	1	1	0	14	
16	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	14	
17	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	14	
45	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1	1	14	
4	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	0	14	
5	1	1	1	1	1	1	1	1	1	0	1	0	0	1	1	1	1	0	14	
6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	15	Upper group (17)
7	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	15	
49	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	15	
50	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	15	
23	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	15	
24	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	15	
10	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	16	
26	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	16	
19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	16	

(continued)

Table 5.3 (continued)

Person	Items in order of difficulty																	
	2	1	5	6.4	3	4	6.3	9.1	9.3	6.1	9.2	6.2	7	10.1	8	10.3	10.2	10.4	R
31	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	16
33	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	16
28	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	17
1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	17
30	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	17
39	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	17
47	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	17
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
Total	48	46	46	46	44	43	43	42	42	41	39	36	36	35	34	24	16	15	
% Correct	96	92	92	92	88	86	86	84	84	82	78	72	72	70	68	48	32	30	

Item Analysis

Range of Difficulties of Items

There does seem to be a range of difficulties of the items even though there are pairs and triplets of items that are of the same or similar difficulty. The item order seems to be close to the order of the items on the test. It would be necessary to consider the content of the items in order to understand the variable. Although items in a proficiency test cannot be expected to be in exactly the same order in the test as their difficulties, they should be close. Similarly, if the test is not supposed to assess the speed of responding, there should be few or no missing responses for items at the end of the test. These and other factors demonstrate the validity of the engagement of the students with the test. Using the Guttman structure as a framework helps in assessing the validity of the engagement.

An Approximation to the Discrimination Index

As mentioned earlier in CTT, the correlation between the scores on an item and the total scores on the test is an indication of the item's discrimination. Here, we study a method for calculating an approximation to the discrimination index that gives a graphical and tangible understanding of the relationship between the responses to an item and the total score. It is termed here simply the discrimination index (DI). Setting up the calculation of the DI leads to a study of the Rasch model for measurement. To calculate the DI, it is necessary to place the persons into groups based on their total scores. Such groups are generally called *class intervals* and we continue to use this term. Here we place persons into just three class intervals, those based on their total scores, the *lower* third (denoted L), the *middle* third (denoted M), and the *upper* third (denoted U). The DI has been approximated in similar ways within CTT with prescriptions for different numbers of class intervals and different proportions of persons in the class intervals. However, because we are calculating the DI here to provide a more tangible understanding of the idea of the discrimination of an item, and to lead into Rasch measurement, and not as an end in itself, we simply use three class intervals where the numbers in each class interval are close to one third of the total sample of persons.

Table 5.3 already shows this classification. Thirteen people are in the lower class interval (denoted NL for the *number* in the *lower* class interval), 20 in the middle class interval (denoted NM), and 17 in the upper class interval (denoted NU). This was the most convenient break-up of the 50 people into three class intervals closely equivalent in size.

To calculate the DI for these items, sum the scores within each of the lower, and upper class intervals, divide these by the total number of persons in the class intervals to form a *proportion* of positive responses in these two class intervals, and then form the difference between these proportions.

This gives the DI as

$$\begin{aligned} \mathrm{DI} &= \mathrm{TSU/NU} - \mathrm{TSL/NL} \\ &= \mathrm{PSU} - \mathrm{PSL} \end{aligned}$$

where

TSU is the *total score* in the *upper* class interval,
TSL is the *total score* in the *lower* class interval,
PSU is the proportion of the maximum score in the upper class interval, and
PSL is the proportion of the maximum score in the lower class interval.

The DI is based on just the data in the lower and upper class intervals. However, we would expect that the proportion of persons who answered an item correctly in the middle class interval would be in between those of the other two class intervals. We use this proportion later to show the discrimination of items graphically. The proportion correct in the middle class interval (PRM) is calculated in the same way:

$$\mathrm{PRM} = \mathrm{NRM/NM}$$

where

PRM proportion right in the middle group,
NRM number right in the middle group, and
NM number in the middle group.

Table 5.4 shows the proportions correct in the three class intervals as well as the index of discrimination (DI) for all the items. The CTT index of discrimination, the correlation (r) between the item and total scores, are also shown in the table for completeness.

Item 2 (the easiest item)

Calculating this index for item 2 gives

$$\begin{aligned} \mathrm{DI} &= \mathrm{PRU} - \mathrm{PRL} \\ &= 17/17 - 12/13 = 1.0 - 0.92 = 0.08 \end{aligned}$$

The value of the DI index is low, very close to 0.0. However, a low value is to be expected with an item that is either very easy or very difficult. If an item is either very easy or very difficult, then most people have the item right or have the item wrong, and therefore the item cannot discriminate among them. *Thus while it is desirable that the item discriminates, if the item is at the beginning of a test and very easy, or at the end of a test and relatively difficult, then a low discrimination should not be used as an indication that there is something necessarily wrong with the item.* Since this is the second item of the test, and therefore expected to be relatively easy, we would not be concerned that it does not discriminate. The CTT correlation index for

Table 5.4 Proportions of persons correct in each of three class intervals and indices of discrimination for all items

Item	2	1	5	6.4	3	4	6.3	9.1	9.3	6.1	9.2	6.2	7	10.1	8	10.3	10.2	10.4
PRL	0.92	0.77	0.77	0.85	0.85	0.62	0.77	0.85	0.77	0.62	0.62	0.38	0.62	0.54	0.31	0.00	0.08	0.00
PRM	0.95	1.00	0.95	0.95	0.85	0.90	0.85	0.70	0.80	0.80	0.70	0.75	0.70	0.65	0.75	0.55	0.35	0.25
PRU	1.00	0.94	1.00	0.94	0.94	1.00	0.94	1.00	0.94	1.00	1.00	0.94	0.82	0.88	0.88	0.76	0.47	0.59
DI	**0.08**	**0.17**	**0.23**	**0.10**	**0.10**	**0.38**	**0.17**	**0.15**	**0.17**	**0.38**	**0.38**	**0.56**	**0.20**	**0.34**	**0.57**	**0.76**	**0.39**	**0.59**
r	**0.25**	**0.36**	**0.33**	**0.06**	**0.05**	**0.52**	**0.24**	**0.20**	**0.29**	**0.37**	**0.42**	**0.43**	**0.31**	**0.41**	**0.48**	**0.60**	**0.42**	**0.53**

this item was 0.25, a low correlation, but not the lowest. Items 3, 6.3, 6.4 and 9.1 had lower correlations.

Let us reflect on the way this index is calculated and why it is a reasonable index to use. First, it is intuitively reasonable to expect that the proportion of the persons answering correctly in the upper class interval should be greater than the proportion answering correctly in the lower class interval. Therefore, this difference should be positive. Further, the greater the difference, the greater the discrimination. Second, since both proportions have to be less than or equal to 1.0, the index must have a number between plus or minus 1.0, as one would require of a correlation. It is empirically possible for a DI and a correlation to be negative, and this would raise serious questions about the item, its scoring and so on.

Item 6.2 (an item of moderate difficulty)

$$\begin{aligned} DI &= PRU - PRL \\ &= 16/17 - 5/13 = 0.94 - 0.38 = 0.56 \end{aligned}$$

This value of 0.56 for the discrimination is greater for item 6.2 than for item 2 which was 0.08. Item 6.2 is of moderate difficulty, and so we should expect it to have a reasonably positive discrimination, above 0.3. The relatively high value of the DI confirms that the item has worked as intended. The CTT correlation index was 0.43.

If an item was of moderate difficulty, and the DI index was close to 0.0, then it would be evident that the item does not discriminate, that is, that the proportion of persons correct on the item in the lower class interval is similar to the proportion correct in the upper class interval. In other words, the proportion correct is not related to the performances on the other items, and therefore the item does not reinforce the information provided by the other items of the test. Therefore, it would be concluded that the item seems to be testing something different from the other items, or that its construction is so bad that it is not working in the expected way. Thus, the item would have to be examined closely to see what has gone wrong with its operation. It is this kind of examination of an item that can be very informative about the construction of the item, about the learning that has taken place, and so on. We will see that this confounding of the discrimination and the difficulty of an item is overcome in Rasch measurement analysis.

Item 7 (a moderately difficult item)

$$\begin{aligned} DI &= PRU - PRL \\ &= 14/17 - 8/13 = 0.82 - 0.62 = 0.20 \end{aligned}$$

The discrimination of this item is not as high as that of item 6.2, but it is still positive. By looking at the pattern of results for these two items in Table 5.4, can you see why the discrimination for item 6.2 is greater than that for item 7? The correlation index for item 7 was 0.31.

Item 10.4 (the most difficult item)

$$\begin{aligned} DI &= PRU - PRL \\ &= 10/17 - 0/13 = 0.59 - 0.00 = 0.59 \end{aligned}$$

In this case, the value of the DI is relatively high. It is high, even though this is the most difficult item because 30% of the persons answered the item correctly. Had only 5% of the persons answered it correctly, the DI would be very small.

Graphical Display of the Item Discrimination

The discrimination of items can be displayed graphically as shown below using the proportions on the vertical axis and the class intervals on the horizontal axis. To locate the scores on the horizontal axis, the *average* score of each of the three class intervals is obtained. This represents the location of each class interval as a whole, and then the proportion of persons who answered the item correctly in each class interval is plotted on the vertical axis at the respective locations of the class intervals.

The plots of these proportions on the vertical axis and the average score of each class interval on the horizontal axis for items 2, 6.2, 7 and 10.4 are shown in Fig. 5.1. It should be evident that item 2 hardly discriminates—the graph is almost horizontal indicating that the proportions correct on the item do not increase with the proficiency of the class interval. However, this is expected because the item is so easy that almost all persons, even those in the lower class interval, answered this item correctly.

We expect that the proportions correct on an item will increase with the proficiency of the class interval. If the proportions correct do increase with the proficiencies of the class intervals, then the item discriminates positively. Items 6.2 and 10.4 discriminate the best, while item 7 has a moderate discrimination.

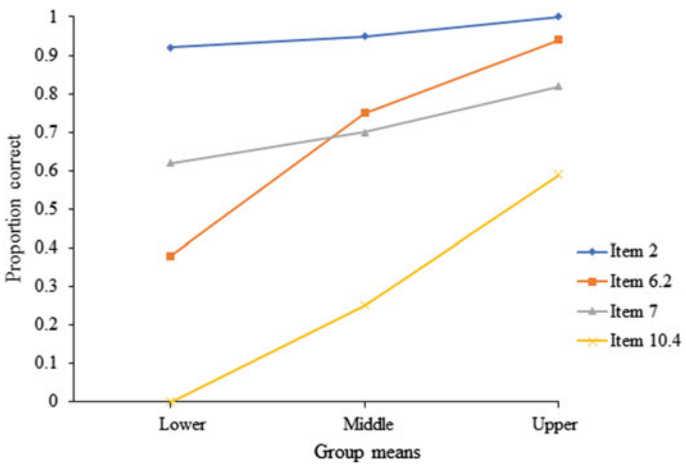

Fig. 5.1 Plots of proportions correct in each of three class intervals for items 2, 6.2, 7 and 10.4

Person Analysis

The person analysis begins at the level of the group of persons. The Guttman structure immediately sets up how many persons obtained a given score, which makes it easy to proceed to make a graph of the frequency distribution. Figure 5.2 shows that the majority of the persons had a total score greater than 10, and that the majority scored 13. It also shows that the distribution has a single mode or peak, with a few people in the tail of the distribution at the lower end. Of course, this is exactly the same information that can be gleaned from Table 5.3 from the columns showing the raw scores and the frequencies in the Guttman table, but it is displayed graphically and reinforces the interpretation.

Overall, in this test and from the perspective of a general context, the distribution shows nothing unusual. One might check who the three persons were that obtained only scores of 6, 7 and 8 respectively, and see what kinds of errors these students made and which items they answered correctly even though these were the easy items. As mentioned earlier in this chapter, in real educational test data, it is most unlikely that the perfect Guttman pattern will be found. We will see that the Rasch model is a *probabilistic* model and requires a probabilistic Guttman structure when items have dichotomous responses. When items are ordered from least difficult to most difficult in the Rasch model, the Guttman response pattern is *the most probable* response pattern for a person.

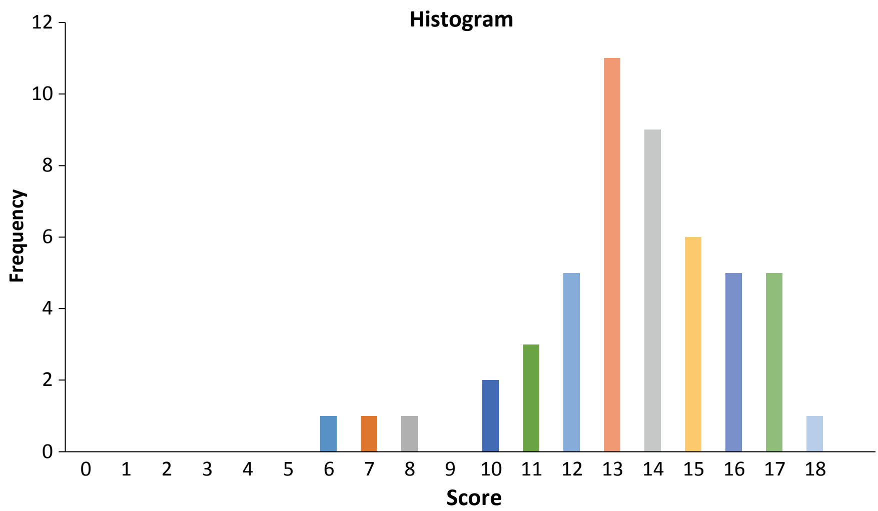

Fig. 5.2 Frequency distribution of scores

Extended Guttman Analysis: Polytomous Items

Tests often contain items that have a maximum score greater than 1, and it might not be possible to break the marks into single marks as easily as was done in Table 5.1. Table 5.5 shows the data rearranged so that scores for items belonging to the same

Table 5.5 Guttman framework including ordered category items

Person	Items in order of total score (difficulty)												
	2	1	5	3	4	7	8	6(4)	9(3)	10(4)	R	F	CF	
38	1	0	1	1	0	0	0	1	2	0	6	1	1	Lower group (13)
2	1	0	1	1	0	0	0	3	1	0	7	1	2	
40	0	1	0	1	1	0	1	3	1	0	8	1	3	
41	1	1	1	1	0	1	0	3	2	0	10	2	5	
42	1	1	0	1	1	0	0	2	3	1	10			
35	1	1	1	1	1	1	0	3	2	0	11	3	8	
44	1	1	1	1	0	1	0	3	3	0	11			
8	1	1	1	0	1	1	0	4	1	1	11			
9	1	1	0	1	1	1	1	2	3	1	12	5	13	
25	1	1	1	0	1	1	0	4	2	1	12			
11	1	0	1	1	1	0	1	3	3	1	12			
46	1	1	1	1	1	1	1	0	3	2	12			
29	1	1	1	1	0	1	0	3	3	1	12			
48	1	1	1	0	1	1	1	4	2	1	13	11	24	Middle group (20)
18	1	1	0	1	1	1	1	4	2	1	13			
32	1	1	1	1	1	1	1	3	2	1	13			
20	1	1	1	0	1	1	1	3	3	1	13			
13	1	1	1	1	1	1	0	2	3	2	13			
22	1	1	1	1	1	1	1	4	1	1	13			
34	1	1	1	1	1	0	1	3	2	2	13			
43	1	1	1	1	1	1	0	4	2	1	13			
36	1	1	1	1	1	1	1	3	2	1	13			
37	1	1	1	1	0	1	1	4	3	0	13			
27	0	1	1	1	0	1	0	4	2	3	13			
12	1	1	1	1	1	1	1	3	0	4	14	9	33	
21	1	1	1	1	1	0	1	3	2	3	14			
14	1	1	1	0	1	1	1	3	2	3	14			
15	1	1	1	1	1	0	1	4	1	3	14			
16	1	1	1	1	1	0	1	4	3	1	14			
17	1	1	1	1	1	1	0	4	3	1	14			

(continued)

Table 5.5 (continued)

Person	Items in order of total score (difficulty)													
	2	1	5	3	4	7	8	6(4)	9(3)	10(4)	R	F	CF	
45	1	1	1	1	1	1	0	3	3	2	14			
4	1	1	1	1	1	0	1	3	3	2	14			
5	1	1	1	1	1	0	1	2	3	3	14			
6	1	1	1	1	1	1	1	4	3	1	15	6	39	Upper group (17)
7	1	1	1	1	1	1	1	4	2	2	15			
49	1	1	1	1	1	0	1	4	3	2	15			
50	1	1	1	1	1	0	1	3	3	3	15			
23	1	1	1	1	1	1	0	4	3	2	15			
24	1	1	1	1	1	1	1	4	3	1	15			
10	1	1	1	1	1	0	0	4	3	4	16	5	44	
26	1	1	1	1	1	1	1	3	3	3	16			
19	1	1	1	1	1	1	1	4	3	2	16			
31	1	1	1	1	1	1	1	4	3	2	16			
33	1	1	1	0	1	1	1	4	3	3	16			
28	1	1	1	1	1	1	1	3	3	4	17	5	49	
1	1	1	1	1	1	1	1	4	3	3	17			
30	1	0	1	1	1	1	1	4	3	4	17			
39	1	1	1	1	1	1	1	4	3	3	17			
47	1	1	1	1	1	1	1	4	3	3	17			
3	1	1	1	1	1	1	1	4	3	4	18	1	50	
Total score	48	46	46	44	43	36	34	166	123	90				
Item	2	1	5	3	4	7	8	6(4)	9(3)	10(4)				
% Maximum	96	92	92	88	86	72	68	83	82	45				

Note R = total score, F = frequency, CF = cumulative frequency

set are added together and persons and items are arranged according to their total scores for a Guttman analysis. It can be seen from Table 5.5 that the totals for items 6, 9 and 10 are greater for these items than for any of the others, but this is in part because the maximum possible score on each item is greater than for the other items. The total scores are 166, 123 and 90 respectively, and these are shown at the bottom of the columns of the respective items.

However, to conduct a Guttman item analysis, it is necessary to take account of the maximum score of these items which is greater than 1. Therefore, in the last row of Table 5.5, the scores are divided by the maximum score of the item, and this is converted to give the % of the maximum possible score. Thus the total for item 6 is divided by 4, that for item 9 by 3 and that for item 10 by 4, because these are the maximum scores that a person could obtain on these items. These figures, 41.5, 41 and 22.5, are now comparable to the scores for the other items. Because these scores are out of 50, they are doubled to convert them to a percentage.

To set up the Guttman structure for the item analysis, we therefore need a second table, called Table 5.6, in which we reorder the items using these new figures. In Table 5.6 we also convert each person's score to a proportion of the maximum, so that it is between 0 and 1. For example, the score of 1 of person 38 on item 6 is converted to 1/4 = 0.25, because the maximum possible score on this item was 4. The persons also must be reordered according to their new total score which is denoted R*. This table is now convenient for further analyzing the persons and items.

Table 5.6 Guttman framework with maximum scores of items converted to 1

Person	Items in order of weighted score (difficulty)											
	2	1	5	3	4	6(4)	9(3)	7	8	10(4)	R	R*	
38	1	0	1	1	0	0.25	0.67	0	0	0.00	6	3.92	Lower group (17)
2	1	0	1	1	0	0.75	0.33	0	0	0.00	7	4.08	
40	0	1	0	1	1	0.75	0.33	0	1	0.00	8	5.08	
42	1	1	0	1	1	0.50	1.00	0	0	0.25	10	5.75	
41	1	1	1	1	0	0.75	0.67	1	0	0.00	10	6.42	
27	0	1	1	1	0	1.00	0.67	1	0	0.75	13	6.42	
8	1	1	1	0	1	1.00	0.33	1	0	0.25	11	6.58	
44	1	1	1	1	0	0.75	1.00	1	0	0.00	11	6.75	
25	1	1	1	0	1	1.00	0.67	1	0	0.25	12	6.92	
11	1	0	1	1	1	0.75	1.00	0	1	0.25	12	7.00	
29	1	1	1	1	0	0.75	1.00	1	0	0.25	12	7.00	
35	1	1	1	1	1	0.75	0.67	1	0	0.00	11	7.42	
9	1	1	0	1	1	0.50	1.00	1	1	0.25	12	7.75	
48	1	1	1	0	1	1.00	0.67	1	1	0.25	13	7.92	
18	1	1	0	1	1	1.00	0.67	1	1	0.25	13	7.92	
34	1	1	1	1	1	0.75	0.67	0	1	0.50	13	7.92	
43	1	1	1	1	1	1.00	0.67	1	0	0.25	13	7.92	
20	1	1	1	0	1	0.75	1.00	1	1	0.25	13	8.00	Middle group (21)
13	1	1	1	1	1	0.50	1.00	1	0	0.50	13	8.00	
37	1	1	1	1	0	1.00	1.00	1	1	0.00	13	8.00	
10	1	1	1	1	1	1.00	1.00	0	0	1.00	16	8.00	
15	1	1	1	1	1	1.00	0.33	0	1	0.75	14	8.08	

(continued)

Table 5.6 (continued)

Person	Items in order of weighted score (difficulty)												
	2	1	5	3	4	6(4)	9(3)	7	8	10(4)	R	R*	
21	1	1	1	1	1	0.75	0.67	0	1	0.75	14	8.17	
14	1	1	1	0	1	0.75	0.67	1	1	0.75	14	8.17	
16	1	1	1	1	1	1.00	1.00	0	1	0.25	14	8.25	
17	1	1	1	1	1	1.00	1.00	1	0	0.25	14	8.25	
45	1	1	1	1	1	0.75	1.00	1	0	0.50	14	8.25	
4	1	1	1	1	1	0.75	1.00	0	1	0.50	14	8.25	
5	1	1	1	1	1	0.50	1.00	0	1	0.75	14	8.25	
46	1	1	1	1	1	0.00	1.00	1	1	0.50	12	8.50	
49	1	1	1	1	1	1.00	1.00	0	1	0.50	15	8.50	
50	1	1	1	1	1	0.75	1.00	0	1	0.75	15	8.50	
23	1	1	1	1	1	1.00	1.00	1	0	0.50	15	8.50	
22	1	1	1	1	1	1.00	0.33	1	1	0.25	13	8.58	
32	1	1	1	1	1	0.75	0.67	1	1	0.25	13	8.67	
36	1	1	1	1	1	0.75	0.67	1	1	0.25	13	8.67	
12	1	1	1	1	1	0.75	0.00	1	1	1.00	14	8.75	
33	1	1	1	0	1	1.00	1.00	1	1	0.75	16	8.75	
30	1	0	1	1	1	1.00	1.00	1	1	1.00	17	9.00	Upper group (12)
7	1	1	1	1	1	1.00	0.67	1	1	0.50	15	9.17	
6	1	1	1	1	1	1.00	1.00	1	1	0.25	15	9.25	
24	1	1	1	1	1	1.00	1.00	1	1	0.25	15	9.25	
26	1	1	1	1	1	0.75	1.00	1	1	0.75	16	9.50	
19	1	1	1	1	1	1.00	1.00	1	1	0.50	16	9.50	
31	1	1	1	1	1	1.00	1.00	1	1	0.50	16	9.50	
28	1	1	1	1	1	0.75	1.00	1	1	1.00	17	9.75	
1	1	1	1	1	1	1.00	1.00	1	1	0.75	17	9.75	
39	1	1	1	1	1	1.00	1.00	1	1	0.75	17	9.75	
47	1	1	1	1	1	1.00	1.00	1	1	0.75	17	9.75	
3	1	1	1	1	1	1.00	1.00	1	1	1.00	18	10.00	
Total score	48	46	46	44	43	41.5	41	36	34	22.5			
Item	2	1	5	3	4	6(4)	9(3)	7	8	10(4)			
% Maximum	96	92	92	88	86	83	82	72	68	45			

Note R = total score, R* = the new weighted total score

Table 5.7 Proportions of persons correct in each of three class intervals and indices of discrimination for items 6, 9 and 10

	Item		
	6	9	10
PRL	0.78	0.71	0.21
PRM	0.8	0.83	0.52
PRU	0.96	0.97	0.67
DI	0.18	0.26	0.46
r	0.51	0.48	0.75

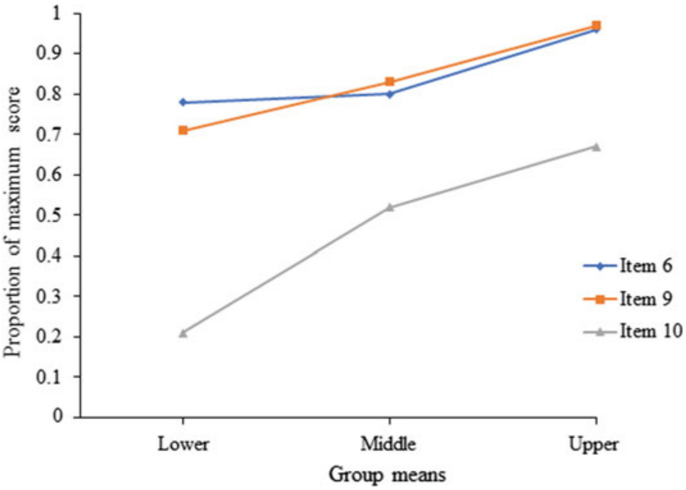

Fig. 5.3 Proportions of persons correct for items 6, 9 and 10

Table 5.7 shows the required proportions to draw the discrimination curves and calculate the discrimination index (DI) for items 6, 9 and 10. It also shows, for completeness, the CTT correlation index.

Figure 5.3 shows a graphical representation of the way the proportions increase as the proficiency of the group increases.

Exercises

1. Organize the data from the Exercises in Chap. 3 in terms of a Guttman structure according to Table 5.5 in this chapter.
2. Draw a frequency distribution of the scores of the persons.
3. Organize the data in terms of a Guttman structure according to Table 5.6.
4. Calculate the discriminations for items 2, 5 and 7.
5. Draw the plots of the discriminations of items 2, 5 and 7 as in Fig. 5.1.

6. Which of these three items is the best discriminating item? Comment on the operation of each item.

References

Bogardus, E. S. (1933). A social distance scale. *Sociology and Social Research, 17,* 265–271.

Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer & others (Eds.), *Measurement and prediction*. New York: Wiley.

Further Reading

Andrich, D. (1985). An elaboration of Guttman scaling with Rasch models for measurement. In N. Brandon-Tuma (Ed.), *Sociological methodology* (pp. 33–80). San Francisco: Jossey-Bass.

Guttman, L. (1950). The problem of attitude and opinion measurement. In S. A. Stouffer & others (Eds.), *Measurement and prediction*. New York: Wiley.