An Introduction to Guttman Scaling

In: Unidimensional Scaling

By: John P. McIver & Edward G. Carmines

Pub. Date: 2011
Access Date: January 29, 2018
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9780803917361
Online ISBN: 9781412986441
DOI: http://dx.doi.org/10.4135/9781412986441
Print pages: 41-61

An Introduction to Guttman Scaling

Historically, Likert scaling developed as a response to the deficiencies attributed to the earlier techniques of Thurstone. Likewise, Guttman scaling, the subject of this chapter, developed as a critical alternative to both of these earlier methods of attitude scaling. In particular, Guttman argued that neither Likert's nor Thurstone's techniques conclusively established that a series of items belong on a *unidimensional* continuum. Evidence that each item is a part of a single underlying dimension, he insisted, is provided by a scale's ability to predict responses to all of its component items on the basis of total scores. This is contrary to the usual intentions of researchers employing item analysis procedures to build a scale; that is, items are chosen for their ability to predict a total score. As we shall discuss in greater detail later, this dispute over the relative merits of the various scaling models largely turns on different conceptions of dimensionality. For now, however, let us introduce Guttman's method of scaling.

Guttman scaling, also known as scalogram analysis and cumulative scaling, is a procedure designed to order *both* items and subjects with respect to some underlying cumulative dimension. It is a deterministic model of scaling; each value is a single-valued function of the underlying continuum (Guttman, 1944: 176).[14] It is in terms of this functional relationship that Guttman defines a scale:

> For a given population of objects, the multivariate frequency distribution of a universe of attributes will be called a *scale* if it is possible to derive from the distribution a quantitative variable with which to characterize the objects such that each attribute is a simple function of that quantitative variable. Such a quantitative variable is called a scale variable [1950: 64].

TABLE 5 A Perfect Guttman Scale

	Items						
Subjects	1	2	3	4	5	6	Scale Score
A	1	1	1	1	1	1	6
B	1	1	1	1	1	0	5
C	1	1	1	1	0	0	4
D	1	1	1	0	0	0	3
E	1	1	0	0	0	0	2
F	1	0	0	0	0	0	1
G	0	0	0	0	0	0	0

A perfect Guttman scale is depicted in Table 5. From each subject's scale score (Guttman's “quantitative” or “scale” variable) we can accurately predict the subject's response to each of the dichotomous items that make up the scale. A score of three, for example, indicates a positive response to items 1, 2, and 3 rather than any other three items. Similarly, a score of one indicates an affirmative response only to the first item. But, if another subject H had answered only item 2 positively, we would no longer have a perfect scale for we cannot accurately predict responses to the individual items given only a knowledge of scale scores. Both F and H would receive scale scores of one even though their responses to individual items differ. These two response patterns cannot be accommodated on the same Guttman scale.

A perfect relationship between scale score and item score (as depicted in Table 5) is rarely if ever achieved. Scalogram *analysis* anticipates that the ideal deterministic model will be violated. The question becomes one of the degree of deviation one is willing to tolerate before it is established that the model fails to serve as an adequate representation of the empirical data. The primary purposes of this chapter are to define deviations from a perfect scale, discuss various criteria for evaluating the degree of deviation, and show how these criteria are used to judge the scalability of a set of items.

Assessment of Error

We define error simply as the deviation of the observed response pattern from the ideal pattern required by the cumulative model. It is assumed that the amount of deviation, or error, observed is strictly a function of the failure of items and subjects to conform to Guttman ordering procedures. Thus, when a set of data exhibits *less* than a specified proportion of errors, it is concluded that the data can be represented on a cumulative scale an underlying dimension, representable by a single quantitative variable. Each item and each subject has a unique meaning in terms of its location on this dimension. However, a wide variety of manipulations exists that make it possible to alter the proportion of error in the data relative to the model.

The two predominant forms of error counting are: minimization of error (Guttman, 1947, 1950; Chilton, 1966, 1969; Green, 1956) and deviation from perfect reproducibility (Goodenough, 1944; Edwards, 1948). A comparison of these two procedures will allow for a more complete investigation of their properties and applications.

According to minimization of error, the number of errors is the least number of positive responses that must be changed to negative or negative responses that must be changed to positive in order for the observed response to be transformed into an ideal response pattern. The results of this procedure are displayed in Table 6, which depicts all possible response patterns and error counts for a four-item Guttman scale.

A contradiction exists between the procedure for counting errors Guttman suggests and the theory of scalogram analysis he develops. The problem makes itself most evident in the following statement:

> Each member of the population will have one of these values assigned to him. This numerical value will be called the person's score. From a person's score we would then know precisely to which problems he knows the answers. *Thus a score of 2 does not mean simply that the person got two questions right, but that he got two particular questions right, namely, the first and second. A person's behavior on the problems is reproducible from his score. More specifically, each question is a simple function of the score* [Guttman, 1950: 66; emphasis added].

The number of positive responses to items is not merely a count of answers given in the affirmative. Rather, it is an assessment of the location of subjects with regard to the proposed cumulative continuum. Consequently, errors as defined by the minimization of error criteria may undercount locational errors by focusing on errors in positive or negative responses.

TABLE 6 Two Forms of Error Assignment for a Four-Item Guttman Scale

	Response Pattern				Assignment of Error	
					Guttman	Edwards-GE
*	+	+	+	+	0	0
*	+	+	+	-	0	0
	+	+	-	+	1	2
*	+	+	-	-	0	0
	+	-	+	+	1	2
	+	-	+	-	1	2
	+	-	-	+	1	2
*	+	-	-	-	0	0
	-	+	+	+	1	2
	-	+	+	-	1	2
	-	+	-	+	2	2
	-	+	-	-	1	2
	-	-	+	+	2	4
	-	-	+	-	1	2
	-	-	-	+	1	2
*	-	-	-	-	0	0

*Indicates ideal scale pattern.

An example will help to clarify this point. If the response pattern (− + + −) is observed, then according to the above statement the fact that the subject scored positively on two items suggests the ideal response pattern (+ + − −). The minimization of error technique described above would count the observed pattern as reflecting one error; the first negative response needs to be changed to a positive one in order to transform the response pattern to the ideal pattern (+ + + −). However, according to the interpretation that Guttman himself places on the scale score, the observed response pattern actually reflects two errors. For if the ideal pattern for two positive items is (+ + − −), and the observed pattern is (− + + −), the minimum number of signs which must be changed to complete the transformation is in fact two. Thus, the cumulative assumption on which a Guttman scale is based is inconsistent with the minimization of error criterion. In such a case, the scale variable loses the most important component of its interpretation.

Guttman fails to address himself to this apparent inconsistency. Edwards (1948), however, derived an error-counting procedure based on the initial assumption that items should in fact be perfectly reproducible from a subject's responses. As Edwards elaborates:

> The suggestion offered here is that we, for a given sample, assume perfect reproducibility, and make our predictions of item responses on this assumption. Error may then be measured in terms of the number of responses departing from the patterns predicted. Cutting points for items would thus be rigorously defined and would always occur between ranks. *Scores which are inconsistent with the assumption of perfect reproducibility would he scored as the nearest scale type consistent with the notion of a scale* [1948:318; emphasis added].

The technique that Edwards asserts is “consistent with the notion of a scale” is exactly that procedure that Guttman referred to when discussing the intended interpretability of two positive responses; assigning ideal response patterns to subjects on the basis of the number of items scored positive (Guttman, 1950: 66). This error-counting procedure, which we will refer to as the Goodenough-Edwards technique, is also presented in Table 6. While error counting based upon deviations from perfect reproducibility will result in a greater number of errors than minimization of error, this is an accurate description of the data based upon scalogram theory. It is for this reason that the deviations from perfect reproducibility method for counting errors is superior to the minimization of error method.

Scale Construction

Let us begin the discussion of scale construction with the simple case of a set of data that conforms perfectly to the cumulative model. The procedures described will also serve us in the more complex case of non-perfect fit discussed shortly. The data matrix of interest is a subject by stimuli arrangement of responses as shown in the top section of Table 7. Note that we are dealing only with dichotomous responses; that is, each subject is permitted only one of two judgments with respect to each stimuli—favorable (1) or unfavorable (0). Focusing on either one or the other response categories (in Table 7, we focus on favorable responses), compute the number of such responses for each individual and the proportion of such responses to each stimuli for all individuals. For example, subject A responds favorably a total of four times and two-thirds of the subjects respond favorably to stimuli 3. Enter these calculations at the margins of the data matrix.

TABLE 7 Construction of a Simple Error-Less Scale

I. Initial Data Matrix

Subjects	Stimuli 1	2	3	4	5	
A	0	1	1	1	1	4
B	0	0	1	0	1	2
C	0	0	0	0	1	1
D	1	1	1	1	1	5
E	0	0	0	0	0	0
F	0	1	1	0	1	3
	.17	.50	.67	.33	.83	

II. Translation of Stimuli (Columns)

Subjects	Stimuli 5	3	2	4	1	
A	1	1	1	1	0	4
B	1	1	0	0	0	2
C	1	0	0	0	0	1
D	1	1	1	1	1	5
E	0	0	0	0	0	0
F	1	1	1	0	0	3
	.83	.67	.50	.33	.17	

III. Translation of Subjects (Rows)

Subjects	Stimuli 5	3	2	4	1	
D	1	1	1	1	1	5
A	1	1	1	1	0	4
F	1	1	1	0	0	3
B	1	1	0	0	0	2
C	1	0	0	0	0	1
E	0	0	0	0	0	0
	.83	.67	.50	.33	.17	

Having accomplished this preliminary task, two more steps are needed. First, interchange the columns of the matrix so that the stimuli are arranged from the highest to lowest proportion of positive responses. This task is shown in the second part of Table 7. Second, arrange the rows of the data matrix so that the subjects are ranked from the greatest number of favorable responses to the fewest. Completion of these two steps (which may be done in either order) will provide the triangular pattern evident in cumulative scales. (See the bottom of Table 7.)

Perfect scales are not often observed. Consequently, we need some means of handling data that do not fit the triangular pattern as well as some measure of the fit between the cumulative model and the empirical data. Two principal techniques of scale construction have been proposed (although there are many variations): the Cornell technique for scale analysis (Guttman, 1947) and the Goodenough-Edwards technique (Edwards, 1957). In each case the error-counting procedure which is employed has been discussed, and it now remains to integrate this procedure with the additional elements of the method.

The Cornell technique for scale analysis is essentially an extension of the minimization of error criterion to a series of observed response patterns. Cutting points are established on the basis of minimization of error, with the only condition being that no item category possess more error than nonerror. The ordering of items on the underlying continuum is likewise a function of minimization of error among the observed responses. As such, the Cornell technique takes advantage of inversions in the ordering of items based upon decreasing marginal probabilities in order to achieve the highest possible reproducibility. The reliance of the Cornell technique on empirical manipulations of data confine the interpretation of scale results. The item ordering and scale scores may be a function of random errors in the data rather than the construct underlying responses to scale items. Consequently, the final scale may be a sample specific result that is neither representative of the population nor replicable in other samples.

In contrast, the Goodenough-Edwards technique for scale analysis is based upon two closely related principles which are consistent with scale theory: (1) the predicted ideal response pattern for a subject is a direct function of the number of items the subject responded to positively and (2) errors are assigned on the assumption of perfect scale reproducibility (Edwards, 1957: 187). Application of these principles insures that cutting points are established objectively and that their placement is consistent with subject response shifts and the resulting rank order.

Beyond its greater theoretical plausibility, the Goodenough-Edwards technique also has some practical advantages. Scale analysis is based on observed responses rather than inferred responses derived under the minimization of error procedure. Consequently, the simple translation of rows and columns described in identifying a perfect scale is appropriate here as a first step prior to error counting.[15]

Following the construction of a proposed scalogram by either method, it is necessary to establish if in fact the scale possesses the theoretically required properties of a cumulative scale.

Every test of scalability is at least partially grounded in the theoretical premise that the observed subject response pattern is a single-valued function of a quantitative scale variable. Since it is accepted that a perfect correspondence between scale variable and observed response pattern is rarely achieved, it is necessary to have a measure that reflects the extent to which the observed response patterns are identical to the predicted ideal response patterns. Guttman originally proposed a statistic, the coefficient of reproducibility (CR), to assess the degree of scalability of empirical data:

> The amount by which a scale deviates from the ideal scale pattern is measured by a

> coefficient of reproducibility. This coefficient is simply a measure of the relative degree with which the obtained multivariate distribution corresponds to the expected multivariate distribution of a perfect scale. It is secured by counting up the number of responses which would have been predicted wrongly for each person on the basis of his scale score, dividing these errors by the total number of responses and subtracting the resulting fraction from one [Guttman, 1950: 77].

The formula for this coefficient may be expressed as follows:

$$\begin{aligned} CR &= 1.0 - (\#\text{ errors})/\text{total responses} \qquad [5] \\ &= 1.0 - (\#\text{ errors})/[(\#\text{ items}) \times (\#\text{ respondents})] \end{aligned}$$

It is obvious that the method of error counting will have a direct affect on the value of CR.

CR, calculated according to the requirements of the particular scale construction technique, is a measure of goodness of fit between the observed and the predicted ideal response patterns. As a result of error considerations, each scaling technique measures CR in a different fashion. Guttman's original formulation will be referred to as CR_{error}, since errors are counted following the successive application of all possible minimizations of error. The CR associated with the Goodenough-Edwards method will be referred to simply as CR_{ge}.

As we have pointed out, scalability is a function of the extent to which observed response patterns can be accurately reproduced on the basis of assigned quantitative scale scores. Guttman established the standard that a set of items should be considered scalable if the observed error in reproduction equals 10% or less of the total responses. However, Guttman's measurement of error was inconsistent with the proposed cumulative interpretation of scalogram theory. As a direct result, CR_{error} fails to reproduce the originally observed response patterns within the stated limits of accuracy (Edwards, 1957: 184).

This does not prove to be the case when error is counted by the Good-enough-Edwards method. With this technique, error is assigned to every observed response which fails to correspond to the ideal scale pattern predicted by the total scale score. The result is that CR, calculated using the Goodenough-Edwards concept of error, accurately reflects the degree to which observed response patterns deviate from ideal response patterns. Retaining Guttman's original specification that a scale is interpretable if it reflects 10% or less error, the scalability criterion now becomes $CR_{ge} \geq .90$. Of course, this is a more conservative test of scalability than that which employs CR_{error}.

The Goodenough-Edwards CR, because of its error-counting procedure, protects against the possibility of a *spuriously* high level of reproducibility. However, it is still true that items with extreme marginal distributions will tend to inflate the value of CR_{ge}. As a safeguard against this possible occurrence, Edwards (1957) suggests comparing CR_{ge} to the minimal marginal reproducibility (MMR).

The calculation of MMR is based on the fact that an item's reproducibility can be no less than the proportion of responses in its modal category.[16] As such, the total reproducibility can be no less than the sum of the proportion of responses in the modal category for each item in the scale, divided by the number of items. This value, MMR, reflects the reproducibility of a series of items *based only upon knowledge of the item marginal distributions.* As an example, consider the marginals for positive responses to four items in Table 8. The marginal probabilities of .8, .6, .4, and .2 (found in row 5) are associated with modal probabilities of .8, .6, .6, and .8. In this case,

$$MMR = (.8 + .6 + .6 + .8)/4 = .7$$

TABLE 8 Comparison of MMR for Various Marginal Distributions of a Four-Item Scale

1	2	3	4	MMR
.50	.50	.50	.50	.50
.60	.60	.40	.40	.60
.70	.50	.50	.30	.60
.70	.70	.30	.30	.70
.80	.60	.40	.20	.70
.80	.80	.20	.20	.80
.90	.70	.30	.10	.80
.85	.85	.15	.15	.85
.90	.90	.10	.10	.90
.95	.85	.15	.05	.90
.95	.95	.05	.05	.95

If the observed marginals were .95, .85, .15, and .05, MMR would equal .90. Thus, the value of MMR is a function of extreme marginals.

The requirements for scalability based upon MMR are as follows: (1) the MMR must not be so large that it is assumed that CR_{ge} is strictly a product of extreme item marginals and (2) the difference between CR_{ge} and MMR must be of such a magnitude that it is possible to attribute an improvement in the prediction of response patterns based upon scalogram analysis.

The relationship between CR and MMR can be illustrated as follows:

$$CR = (TR - SE)/TR \quad [6]$$

$$MMR = (TR - ME)/TR \quad [7]$$

where

TR = total responses
SE = scale errors
ME = marginal errors, the sum of all nonmodal frequencies.

The difference between the two coefficients is

$$CR - MMR = [(TR - SE)/TR] - [(TR - ME)/TR] \quad [8]$$
$$= (ME - SE)/TR$$

In words, the difference between these coefficients is a function of the improvement in prediction provided by the scale over the marginal frequencies of individual items. This difference ranges from 0 (if the scale provides no improvement of prediction) to ME/TR (if the scale fits the Guttman criteria perfectly). Given that the maximum marginal errors in any item that can occur are 50%, this difference has a theoretical maximum of .50. The maximum that will occure in any sample is a function of the item marginals.

Interpreting the difference between CR and MMR on a scale from 0 to ME/TR can be difficult. As a consequence, various alternatives have been suggested. For example, Menzel (1953) also noticed that CR would be necessarily high if item marginals were substantially skewed regardless of any relationship among the items. He argued that the number of scale errors attributable to each item cannot be higher than the frequency of responses to the nonmodal category. Consequently, the maximum number of scale errors is the sum of the responses to the nonmodal category of each item. Menzel designed the coefficient of scalability (CS) as a measure of a scale's ability to predict item responses in comparison to predictions based on marginal frequencies. The formula for CS is:

$$CS = 1.0 - (\text{scale errors}/\text{marginal errors}) \quad [9]$$

Todd (1974) suggests that CS is easily interpretable as a proportional reduction in error (PRE) statistic[17]:

$$CS = (ME - SE)/ME \quad [10]$$

As a measure of improvement in fit, CS provides fixed reference points. If scale predictions are perfect (i.e., there are no scale errors), CS = ME/ME = 1.0. If the scale provides no improvement in prediction (i.e., scale errors equal marginal errors), CS = 0. Menzel suggests a coefficient of .60 is an indication of scalability. This is, however, only a rule of thumb; a CS of .60 has no explicit theoretical justification.

The final criterion for scalability involves analysis to insure against the presence of a second scale variable (i.e., a second underlying scalable dimension). Setting the minimum acceptable value of the coefficient of reproducibility at .90 means that 10% error can be tolerated while the data are considered scalable. Scalogram analysis, however, assumes the presence of only one underlying scalable dimension. If a particular *error* response pattern is observed frequently in the data, this is an indication that the data should be represented by more than a single dimension (Guttman, 1947: 457). Thus, in order to establish that a set of items is scalable, it is necessary that observed error responses be essentially random, thereby providing no evidence of a second underlying dimension.

In summary, the Goodenough-Edwards technique for scale analysis stipulates three necessary criteria which must be fulfilled in order that a series of ordered items be considered scalable. The first, similar to Guttman's procedure, is that the coefficient of reproducibility, CR_{ge}, be greater than or equal to .90. Second, the MMR must not be excessively high, and the difference between CR_{ge} and MMR must indicate that some improvement in scalability is realized as a function of knowledge of total scores. And finally, those response patterns which reflect error must be non-systematic in character.

Assignment of Scale Scores

The final step in scale analysis is the assignment of scale scores to subjects. The assignment of scale scores would be straightforward if only perfect scales were observed; the quantitative variable would be equal to the number of positive responses. For example, an observed response pattern (+ + + − −) is assigned the scale variable "3." The number 3 allows the inference that the subject not only responded positively to three items, but that these three items were *the first three.*

The problem arises when nonscale responses are assigned ideal predicted response patterns and scale scores are assigned according to these ideal patterns. If the response pattern (− + + − −) is observed, there is no number from 0 to 5 which accurately locates the respondent on the underlying scale continuum. However, nonscale responses do occur, and the researcher is faced with the problem of what score to assign so that interpretation of the scale will be least affected. This assignment problem, as we shall see, is in fact a pseudo-problem, based upon minimization of error scale construction techniques.

Two predominant forms of scale score assignment have been developed for scales constructed through minimization of error techniques: (1) assignment based upon nearest ideal predicted response pattern and (2) assignment based upon the number of positive responses. To differentiate between these procedures, assume that for a five-item scalogram, one response pattern is observed to be (− + + − −). Based upon the minimization of error assessment of nearest ideal scale type, the response pattern would be reclassified as the pattern (+ + + − −)

and assigned the scale variable 3. However, because there are only two pluses in the original observed pattern, the number of positive responses method assigns the scale variable 2.

This lack of correspondence between observed response pattern and either scale score procedure introduces a serious restriction on the interpretation of scale results for subjects who display nonscale responses. Equally as damaging is the case where an observed response pattern could technically be assigned to two or more ideal predicted types. The researcher, in trying to decide which ideal pattern the observed response is actually a deviation from, is forced to make an ambiguous assignment as to how the subject should be characterized with regard to the scale attribute. For the observed response pattern (− + − + −), the ideal predicted response patterns based upon minimization of error could be either (+ + + + −), (+ + − − −), or (− − − − −). Scale score assignment based upon nearest ideal type is either 4, 2, or 0, depending upon the choice of the researcher. Scale score assignment based upon the number of positive responses observed in the original pattern is again equal to 2.

The assignment of scale score problem illustrated above is a pseudo-problem once the presence of a scalogram is established by the recommended criterion that CR_{ge} be greater than or equal to .9. When a scale has been constructed in this manner, rather than by minimization of error techniques, the assignment of scale scores based on the nearest ideal type is identical to the assignment of scale scores in accordance with the number of positive responses. The assignment of scale scores to subjects is not an issue when the Goodenough-Edwards procedures have been employed, because the number of positive responses is identical to the nearest ideal scale type. This method of score assignment is also much easier to use. In summary, scale construction and scoring methods based on the perfect reproducibility model is preferred to the minimization of error approach.

The calculation of scale scores for the cumulative scale by summing the number of positive responses sounds as if it is the same scoring rule used in summative or Likert scales. In fact, the assignment procedure is identical. The distinction between the cumulative and summative models rests in *when* the responses are totaled and how this total is interpreted. Scale scores are assigned to respondents only if all (or at least a substantial subgroup) of their responses fit the model being considered. Guttman scale scores are assigned only when CR is greater than .90, for example. Likert scores are computed only if the item-to-total correlations are statistically significant and greater than the interitem correlations. It is certainly possible for data to fit one model and not the other or that neither model is appropriate.

The interpretation of total scores distinguishes these two scaling models. A Likert score of 2, for example, means that the respondent replied favorably to *any* 2 of n stimuli. An individual who responds positively to two items out of n that compose a Guttman scale has responded to two *specific* items, the two that are the "easiest" or "most acceptable" to the group of respondents.

Guttman Scaling and Item Analysis

Edwards (1957: 172) argues that scalogram analysis is not strictly a method for constructing or developing an attitude scale. Rather, it is a process by which it is determined whether a series of items and a sample of subjects conform to a specified set of criteria designated as the requirements of a Guttman scale. To this point we have treated Guttman scaling from this perspective. That is, Guttman scaling has been discussed from a hypothesis-testing standpoint.

Yet the cumulative scaling model has also been used as an exploratory technique for selecting items that conform to scale criteria from a larger set of items. This selection process occurs quite frequently and is in large part a function of our inability to establish the homogeneous content of a series of items prior to scaling. One method of item selection that has proved useful is the examination of bivariate relationships among all items using a correlational model that is consistent with the cumulative assumption of Guttman scaling.

The bivariate relationships between responses to the components of a four-item Guttman scale are shown in Table 9. (Here we are dealing with a perfect scale.) The four items partition the population into five groups, those individuals that do not accept any items (n_5), those that accept only item 1 (n_4), and so forth. In order to be consistent with the cumulative model, the relationship between any two items must be a weak monotonic relationship. That is, low responses may be paired; high responses may be paired; a high response on one variable may be paired with a low response on another, but not vice versa. Those conditions are maintained in each of the 2 × 2 components of the large matrix of relationships in Table 9.

What correlation coefficients are appropriate measures of weak monotonic relationships? Several are available. Torgerson (1958) suggests tetrachoric r. Robinson et al. (1968) offer Yule's Y as an alternative. Yule's Q has perhaps been most frequently used. MacRae (1970) has demonstrated its considerable value in the study of congressional voting by analyzing large numbers of roll call votes. All of these coefficients reach maximum values of 1.0 under conditions of weak monotonicity.[18] If the relationships are not weakly monotonic, the coefficients will all have different values with Q being the largest. Magnitude of relationship as a selection criterion has largely been ad hoc. MacRae and others who use this technique examine alternative cutoffs in attempting to find the most theoretically plausible cluster of items. In other words, ad hoc cutoffs exist to serve the researchers' purposes. Cutoffs of .9 or .8 for

Yule's Q are most common.

TABLE 9 Item Analysis and Guttman Scaling*

0000	I ↓	1000	II ↓	1100	III ↓	1110	IV ↓	1111	- response pattern
n_5		n_4		n_3		n_2		n_1	- population

Item 4 is accepted by n_1 respondents.
Item 3 is accepted by n_1+n_2 respondents.
Item 2 is accepted by $n_1+n_2+n_3$ respondents.
Item 1 is accepted by $n_1+n_2+n_3+n_4$ respondents.

I \ II	0	1
0	n_5	0
1	n_4	$n_1+n_2+n_3$

II \ III	0	1
0	n_4+n_5	0
1	n_3	n_1+n_2

I \ III	0	1
0	n_5	0
1	n_3+n_4	n_1+n_5

III \ IV	0	1
0	$n_3+n_4+n_5$	0
1	n_2	n_1

II \ IV	0	1
0	n_4+n_5	0
1	n_2+n_3	n_1

I \ IV	0	1
0	n_5	0
1	$n_2+n_3+n_4$	n_1

Q, Y, Tetrachoric r

I	1.0			
II	1.0	1.0		
III	1.0	1.0	1.0	
IV	1.0	1.0	1.0	1.0

*Adapted from Torgerson, 1958.

Senate Voting—The Structure of Support for a Consumer Protection Agency (An Example of Guttman Scaling)

In 1975, the U.S. Senate made its fourth attempt in 5 years to establish a federal consumer protection agency (Agency for Consumer Advocacy; ACA) to represent consumer interests before other federal agencies and courts. As conceived by its sponsors, the new agency would not be able to issue regulations on its own, but would petition other federal agencies to do so. The ACA would also be charged with gathering information of interest to consumers.

Fourteen roll call votes were taken on S200 and proposed amendments (see Table 10). Sponsors were able to turn back most of the attempts to weaken the legislation with one exception. An amendment (offered by Senator Dole) to prevent the ACA from participating in federal proceedings on agricultural issues was accepted by a 2–1 margin (CQ181).

The question we pose is whether the Senators were expressing some unidimensional attitude in voting on the amendments to S200. In asking the question this way, we initially eliminate the two votes CQ171 and CQ184, the vote on cloture and the final passage vote. While these two votes fit our final scale, they were removed from our analysis to ease interpretation of the results of our scaling efforts. Thus, our set of roll calls for analysis are the 12 substantive amendments to S200. Opponents of this legislation attempted to amend the bill to include many exceptions. Consequently, a final scale might be interpreted as a ranking of roll calls supporting special interests. Alternatively, the liberal position is support for the original bill, that is, for a viable and general consumer protection agency.

The correlations among the 14 roll call votes are presented in Table 11. We can see immediately that CQ169 (an FCC exception) does not appear related to the rest of the roll call votes. In addition, CQ181 (the Dole Amendment exempting farm issues) is not related to a number of the other votes. These same conclusions may be drawn from an initial Guttman scale of these 12 roll calls. CQ169 and CQ181 contribute a majority of the scale errors (38 of 104 or 37%). Finally, the scale does not meet our evaluative criteria (CR =. 874).

TABLE 10 Senate Roll Call Votes on S200, a Bill to Establish an Agency for Consumer Advocacy (ACA)—May 1975

CQ169 Magnuson amendment to add a provision that would prohibit the ACA from intervening in the broadcast license proceedings of the FCC (Adopted 69-21).

CQ170 Weicker amendment to delete a provision in the bill that would prevent the ACA from intervening in labor-management disputes (Rejected 37-51).

CQ171 Ribicoff motion to close further debate (Adopted 71-27).

CQ174 Johnson amendment to limit the ACA's authority to seek judicial review of other agencies' actions to cases that affect consumer health or safety (Rejected 29-63).

CQ175 Helms amendment to require the ACA to notify everyone against whom a consumer complaint has been lodged and that these complaints be made public (Rejected 41-50).

CQ176 Pearson amendment, in the form of a substitute bill, to set up an Office of Consumer Counsel in each of 24 major federal departments and agencies (Rejected 22-70).

CQ177 Taft amendment to permit the ACA to intervene in NLRB proceedings dealing with secondary boycotts and jurisdictional strikes (Rejected 36-57).

CQ178 Ribicoff motion to table the McClure amendment to prohibit the ACA from intervening in federal procedings with the intention of limiting the sale, manufacture, or possession of firearms (Rejected 27-67).

CQ179 Taff amendment, in the form of a substitute bill, to set up an Office of Consumer Affairs in the executive branch and corresponding offices in other federal departments and agencies (Rejected 28-64).

CQ180 McClellan amendment to replace the single ACA administrator with a three-member commission appointed by the President with Senate Confirmation (Rejected 40-48).

CQ181 Dole amendment to prevent the ACA from intervening in federal proceedings directly affecting producers of livestock, poultry, or agricultural commodities (Adopted 55-34).

CQ182 Scott amendment to authorize the Justice Department to control ACA litigation instead of leaving it to ACA's attorneys (Rejected 24-67).

CQ183 Griffin amendment, in the form of a substitute bill, to delete all provisions except the requirement that federal agencies prepare cost-benefit assessments on proposed regulations (Rejected 24-66).

CQ184 Passage of S200, a bill to set up an independent Agency Consumer Advocacy to represent consumer interests before other federal agencies and courts, and to gather and disseminate consumer information (Passed 61-28).

TABLE 11 Correlations Among Roll Call Votes (Yule's Q)

CQ#														
169	100													
170	79	100												
171	19	93	100											
174	05	.93	99	100										
175	21	90	100	100	100									
176	-26	85	94	98	94	100								
177	75	100	96	95	91	84	100							
178	-08	77	100	100	95	83	79	100						
179	14	95	99	100	100	97	97	100	100					
180	15	83	100	98	94	96	85	94	97	100				
181	-51	93	91	86	69	77	31	70	84	83	100			
182	24	89	98	98	100	92	88	86	97	93	59	100		
183	19	93	100	99	100	89	97	100	99	100	91	97	100	
184	19	91	99	98	100	91	94	100	100	100	84	86	100	100

A scale of the remaining 10 roll call votes provides an acceptable scale by most criteria (CR = .923, MMR = .645, CS = .782). Furthermore, no error patterns predominate. However, the data permit us to illustrate the construction of a “contrived” item, that is, a combination of responses to a subset of items. The roll calls on CQ170, 177, and 180 all fall at approximately the same point on the scale. This is also the location of many of the scale errors. We might argue that there are no real differences between these roll calls and that inconsistent votes are random errors that should be ignored. We combine items as follows: one or fewer positive responses are coded as a no response to the contrived item; two or more positive responses are coded as a yes response. The combination of two of these items (CQ170, 177) also makes some intuitive sense in that they deal with a common exception—labor-management disputes.

The final scale (Table 12) is a minimal improvement over the 10-vote scale (CR = .9433, MMR = .6650, CS = .8308). Over 78% of the Senators fit one of the nine perfect scale patterns. The exemption most supported by the Senators was denial of control over firearms. The alternative least supported was Senator Pearson's alternative that the ACA be scrapped for internal consumer advocates in each of the federal departments.

There is an epilog: Advocates of the bill did not rejoice at the passage of S200 for long. While the House approved the creation of a consumer protection agency late in the year, the slim margin of victory (208–199) disappointed supporters of this legislation. With so little support in the House and under the threat of a veto from President Ford, the sponsors in both houses decided against convening a conference to consider the differences between the House and the Senate bills.

TABLE 12 A Guttman Scale of Senate Voting

CQ vote	178	135	180 187 190	184	189	192	193	186	
scale score									
8	1	1	1	1	1	1	1	1	- Case, Hart (P), Humphrey, Inouye, Jackson, Pastore, Proxmire, Ribicoff Williams, Mondale, Nelson, Pell, Brooke, Percy, Cranston, Stevenson, Glenn, Hart (G)
7	0	1	1	1	1	1	1	1	- Bentsen, Burdick, Cannon, Church, Magnuson, Mansfield, Moss, Muskie, Scott (H), Schweiker, McIntyre, Eagleton, Clark, Huddleston, Leahy
6	0	0	1	1	1	1	1	1	- Byrd (R), Montoya, Randolph, Stafford, Abourezk
5	0	0	0	1	1	1	1	1	- Roth, Chiles, Stone
4	0	0	0	0	1	1	1	1	- Domenici
3	0	0	0	0	0	1	1	1	
2	0	0	0	0	0	0	1	1	
1	0	0	0	0	0	0	0	1	- McClelland, Byrd (H), Scott (P), Allen, Helms
0	0	0	0	0	0	0	0	0	- Curtis, Eastland, Griffin, Hruska, Thurmond, Young, Fannin, Tower, McClure, Hansen, Bartlett, Garn
Error Patterns									
7	1	1	0	1	1	1	1	1	- Hathaway, Weicker
7	1	1	1	1	1	1	1	0	- Pearson
6	0	1	0	1	1	1	1	1	- Hatfield, Hollings, Haskell
5	1	0	0	1	1	0	1	1	- Fong
4	0	0	1	1	1	0	1	1	- Packwood
4	0	0	0	1	1	1	0	1	- Sparkman
3	0	0	1	0	1	0	1	0	- Stevens
2	0	0	0	0	0	1	0	1	- Stennis
2	0	0	0	0	0	1	1	0	- Dole, Brock
1	0	0	1	0	0	0	0	0	- Talmadge
1	0	0	0	0	0	0	1	0	- Taft
1	0	0	0	0	0	1	0	0	- Nunn
% Support	29	52	56	67	69	71	72	75	
# Errors	4	3	8	1	1	7	3	7	

Participation in Politics—A Cumulative Scale?

Students of politics have noted that persons who engage in one political act often participate in other acts. Milbrath (1965) argues that involvement in politics is structured in a certain way: Participation is hierarchical. This hierarchy of political acts is organized such that individuals who perform acts at the top of the hierarchy also perform acts lower in the hierarchy. Individuals at the bottom of the hierarchy, however, perform only the lowest ranking (most frequent) political acts.

The logic underlying this argument goes something like the following. Political participation may be ranked hierarchically in terms of the personal commitment (defined as time, energy, and money) required for each act. The apathetic citizen invests little in the political world and receives little in return. Voting is routine. It is the easiest way for many citizens to act politically. Many other political activities, however, are available to citizens. They may discuss politics with family and friends. This discussion can be carried to the point of attempting to persuade others to accept a point of view. Citizens can express their opinions publicly. This may be done passively by displaying campaign buttons and bumper stickers or actively by discussion with strangers. Citizens can invest more time and effort in politics by contacting local and national public officials, contributing to campaigns, and attending political meetings and rallies. Finally, citizens may participate by working in campaigns, soliciting funds, or even running for and holding public office.

Milbrath's hypothesis that political participation is hierarchically structured is equivalent to the hypothesis that political acts would conform to a Guttman scale. This scale of political acts is portrayed in Table 13. Initial empirical support for this scale consisted of aggregate frequencies of occurrence. Voting is the most frequent political act. Membership in a political party is the least common one.

The cumulative scale in Figure 1 implies more than the ranking of actions by aggregate proportions of the populations so engaged. The scale implies particular constraints on *individual* behavior. Citizens who joined political parties have also performed all other political acts. If citizens perform only one political act it is voting. Individuals involved in four political acts participate in four *particular* acts—no action higher on the hierarchy than wearing a campaign button will be undertaken.

Does individual behavior fit the constraints of the cumulative scale? Verba and Nie (1972) suggest that participation by American citizens is not hierarchically structured. Empirical evidence provided by a national survey of political behavior suggests that the Milbrath hypothesis is not supported:

TABLE 13 Milbrath's Hierarchy of Political Involvement—A Hypothesized Guttman Scale

Citizens	Vote-Natl Elections	Vote-Local Elections	Attempt to Persuade Others	Wear a Campaign Button	Attend Rallies	Contact Public Officials	Contribute Money	Join Political Party	Scale Score
Politically Active (Hi)	1	1	1	1	1	1	1	1	8
	1	1	1	1	1	1	1	0	7
	1	1	1	1	1	1	0	0	6
	1	1	1	1	1	0	0	0	5
	1	1	1	1	0	0	0	0	4
	1	1	1	0	0	0	0	0	3
	1	1	0	0	0	0	0	0	2
	1	0	0	0	0	0	0	0	1
Politically Apathetic (Lo)	0	0	0	0	0	0	0	0	0

> Researchers simply may have overestimated the degree of structure in and the amount of correlation among the varieties of participatory acts when interpreting simple frequency distributions [Verba and Nie, 1972: 40].

Verba and Nie discovered that 53% of their respondents participated in none of the six most difficult political acts. If these data were to fit the cumulative model, however, 80% of the population should not have performed any "difficult" political acts (as only 20% of the population participated in the "easiest" of the six political acts). Political activity is not cumulative, according to Verba and Nie, but is a multidimensional phenomenon with several factors influencing the participatory behavior of American citizens.

Congressional voting and political participation are just two subjects where cumulative scaling have been applied. Guttman's procedures have been used extensively in the construction of attitude indices by social scientists (see D. Miller, 1970, and Robinson et al., 1968). But while the technique was designed as a means to examine the dimensionality of a series of qualitative responses to attitude surveys, scaling has been used by many social scientists to answer a variety of research questions. Anthropologists have employed Guttman scaling in their studies of cultural evolution (Carneiro, 1962) and development in rural communities (Young and Young, 1962). Sociologists have applied analysis to the study of bureaucratic structures (Udy, 1958), leisure activities (Allardt et al., 1959), evolution of legal institutions (Schwarts and Miller, 1962), sexual experiences (Podell and Perkins, 1957), and neighboring activities (Wallin, 1953). Political scientists have found cumulative scaling a useful technique for examining political development (Snow, 1966), international conflict and cooperation (Moses et al., 1967), political violence (Nesvold, 1971), and congressional (e.g., Rieselbach, 1966; MacRae, 1970; Clausen, 1973; Weisberg, 1974) and judicial decision making (e.g., Schubert, 1968; Spaeth, 1969).

In this chapter we have provided a brief overview of Guttman scaling. The simplicity with which we have described this procedure is, unfortunately, not an adequate summary of the technique. Almost all aspects of Guttman scaling from its goodness of fit tests to its conclusiveness as a measure of unidimensionality have been subjected to many criticisms, some of which we will discuss in the next chapter.

http://dx.doi.org/10.4135/9781412986441.n4