

**SHORT TAKE**

## **Using Guttman Scaling to Rank Wealth: Integrating Quantitative and Qualitative Data**

GREG GUEST  
*Sapient Corporation*

*Wealth ranking in given field sites can be problematic for a number of reasons. This article explores the usefulness of Guttman Scaling and AnthroPac software in such contexts, using a small fishing community on the northern coast of Ecuador as an example. The author provides a step-by-step description of procedures for implementing and analyzing Guttman Scale methodology and discusses the issue of construct validity. The complementary relationship between qualitative and quantitative data is highlighted throughout.*

**A**nthroPac (Borgatti 1992) is a suite of programs for collecting and analyzing a variety of field data. In this article, I show (1) how to use the routine in AnthroPac to build a Guttman scale for assessing household wealth and (2) how a Guttman scale, informed by qualitative data, can provide insights and generate interesting questions for further study. Finally, I discuss some of the potential problems and questions that arise when using the technique.

Household wealth is a widely used variable in social science research, but for researchers working in developing countries, where income is not documented or wage labor is uncommon, assessing wealth can be a daunting task. One potential solution is wealth ranking: Individuals or households are ranked relative to each other according to given criteria and subsequently given an ordinal value. Grandin (1988) had local participants rank families

---

*The Short Takes section of Field Methods is for brief research articles and how-to articles like the one by Greg Guest in this issue.*

*Many thanks to Jeffrey Johnson, Mike Burton, and all of the instructors at the NSF Summer Institute for Methods and Research Design in Cultural Anthropology. My gratitude also goes to the kind people of Ecuador for their friendship and patience and to the Social Science and Humanities Research Council of Canada that funded my research.*



*Field Methods*, Vol. 12, No. 4, November 2000 346–357  
© 2000 Sage Publications, Inc.

into different groups according to perceived wealth; Adams et al. (1997) found that this method produces a reliable and valid measure but that getting people to participate is difficult. Another option is to do household inventories, calculating the total value of goods in a sample of houses. This is a very time-consuming task, and often people don't remember the prices they paid for items in their homes. A convenient alternative is Guttman scaling.

### GUTTMAN SCALING

Developed in 1944 (Guttman 1944), the Guttman Scale measures the extent to which a series of items is distributed unidimensionally. Essentially, two questions are asked: (1) If a unit of analysis (a person, a household, an organization, a country, etc.) exhibits some trait, then does that unit of analysis have certain other traits as well? (2) Is there a particular *order* in which these traits are accumulated or manifested? Carneiro (1962, 1973) used Guttman scaling to assess the distribution of cultural traits across societies (if an ancient society had paved streets, for example, then what other traits did it have?). Werner (1985) used Guttman scaling to develop an index of stress for Brazilian farmers. In another article, Johnson (1998) tested whether the manufacturing skills of Matsigenka men (the Matsigenka are a group of Amazonian Indians in Peru) are unidimensional. That is, if a Matsigenka man knows how to make a canoe, then can we predict accurately what other skills he has?

Guttman scaling has been used by anthropologists to assess status and wealth in small communities (Hammel 1962; Kay 1964; Pelto 1973; Pelto and Pelto 1978; DeWalt 1979:108–10), and the method has a lot of appeal as a tool for field assessment of wealth. But in the past, the analyses were done by hand, and data manipulation was labor intensive. AnthroPac's Guttman routine (Borgatti 1992) makes the analysis quick and easy.

### PROCEDURE

I was working with fishermen in a small village on the coast of Ecuador and needed a quantitative measure of wealth. The goal was to establish an ordinal value for each household and to use the value in a variety of statistical analyses. Capital flow is variable and unpredictable in this community—few people are regularly employed—so asking direct questions about income would have been futile. Comprehensive household inventories were too labor intensive and time consuming—for informants as well as for me.

I started by asking key informants about how people differ with respect to wealth and how these differences are manifested. I asked people to talk about levels of wealth (or more accurately, levels of poverty) in the village, and what separated, for example, the *pobres* (poor) from the *bien pobres* (really poor) or the *bien, bien pobres* (really, really poor).

After several conversations, it became clear that a family's wealth in this community is measured according to their material possessions. Locals told me, for example, that the really poor do not have a gas stove and that the really, really poor cannot even afford mosquito nets. Proceeding like this with a number of different informants, I developed a list of six items that locals saw as salient in differentiating people in terms of wealth.

1. mosquito net
2. mattress
3. gas stove
4. television
5. refrigerator
6. stereo

Although these six items might have been enough, I added six more items, based on my own observations and from interviews. After all, items can be deleted, but never added, after data collection.

7. electricity
8. savings account
9. boat
10. outboard motor
11. land
12. vehicle

I included the full list of twelve items, in a simple yes/no format, on a survey administered to 203 households in the village, or every second house. That is, for each house, I asked whether each of the twelve items was present or absent. I entered these data into a respondent-by-item matrix in AnthroPac (it took me about ninety minutes to type in the 203 households and twelve items), with each row representing a household and the twelve columns representing the items. Entries in the original data were ones and zeros, indicating that a household either had or did not have a particular item. Table 1 shows a portion of the output produced by the Guttman scaling routine in AnthroPac. Note that as the list moves downward, households have progressively fewer items.

TABLE I  
The Guttman Scale with Twelve Items (a partial sample)

ID#	M	E	Ma	S	T	L	R	So	B	Mo	Sa	V	Score
115	1	1	1	1	1	1	1	1	1	1	1		11
051	1	1	1	1	1	1	1	1	1	1			10
038	1	1	1	1	1	1	1	1	1	1			10
129	1	1	1	1	1	1	1	1	—		+		9
080	1	1	1	1	1	1	1	—	1	+			9
110	1	1	1	1	1	—	1	—	1	+	+		9
022	1	1	1	1	1	1	1	—	1	+			9
109	1	1	1	1	1	1	1	—	1	+			9
160	1	1	—	1	1	—	1	1	1	+	+		9
036	1	1	1	1	1	1	1	1	—		+		9
057	1	1	1	1	1	1	1	1	—		+		9
076	1	1	1	1	1	1	1	—	1	+			9
023	1	1	1	1	1	1	1	—	1	+			9
013	1	1	1	1	1	1	—	1	1	+			9
093	1	1	1	1	1	1	1	1					8
074	1	1	1	1	1	1	1	1	1				8
141	1	1	1	1	1	1	1	1	—			+	8
117	1	1	1	1	1	—	1	1			+		8
046	1	1	1	1	1	1	1	1					8
069	1	1	1	—	1	1	1	—	+	+			8
001	1	1	1	1	1	1	1	1					8
003	1	1	1	1	1	1	1	—	+				8
092	1	1	1	1	—	1	—	—	+	+	+		8
134	1	1	1	1	1	1	1	1					8
100	1	1	1	1	1	1	1	1					8

NOTE: N = 203. Total number of errors = 292; coefficient of reproducibility = 0.880. ID# = identification number; M = mosquito net; E = electricity; Ma = mattress; S = stove; T = television; L = land; R = refrigerator; So = stereo; B = boat; Mo = motor; Sa = savings account; V = vehicle.

The number in the left column refers to the respondent identification. The adjacent row shows which of the twelve items a household has, indicated by a one or by a plus sign. The last column is the total number of items in each household—that is, the sum of the 1s and the +s. The plus signs, however, indicate errors of inclusion. These occur when a household has an item higher in the scale but does not have all the items lower in the scale. For example, household 129 has a plus sign for “savings account” (column 11), indicating an error of inclusion. If a household has a savings account, then it should also have an outboard motor and a boat since the pattern of responses place those

items earlier on the scale. Household 129 has neither a boat nor an outboard motor, so its savings account counts as an error of inclusion.

The minus signs indicate errors of omission. These occur when a respondent fails to possess a particular item, given that it has items higher in the scale. Household 80 shows an error of omission for “stereo” (column 8). This house does not have a stereo but does own a boat and motor, which are higher up the scale. The lack of a stereo is counted as an error of omission.

Errors of inclusion and omission in a Guttman Scale can be counted in two ways. The first, developed by Guttman (1950) himself, is called *minimization of error* and is measured by counting the least number of responses that must be changed to transform a respondent’s response pattern into an ideal scale. In this procedure, an ideal scale refers to the ordering of items and does not consider the total number of items a respondent may have. The second method, *deviation from perfect reproducibility*, is more conservative (i.e., it will tend to calculate more errors) and assigns errors based on an ideal response pattern as measured by the order of responses *and* the total number of items a respondent possesses (Goodenough 1944; Edwards 1948). AnthroPac uses this latter, more conservative, method.

The difference between these two measures can best be illustrated by example. Referring to Table 1, notice that household 110 has four errors. This household actually has nine items. To achieve a perfect scale based on having the particular array of nine items for household 110, we would need to change the two errors of omission in columns 6 and 8 to ones. This would give household 110 a total of eleven items, so we would also have to drop the items in columns 10 and 11. This would yield a perfect scale of nine items. In the less conservative method, all we would need to do to create a perfect scale is change the negative scores in columns 6 and 8 to positives. This would produce only two errors, but a total of eleven items for the household when, in fact, it only had nine.

AnthroPac also calculates Guttman’s Coefficient of Reproducibility, or CR.

$$CR = 1 - \text{number of errors/number of entries}$$

The CR is a measure of the unidimensionality of the items in scale and by convention, a CR of .90 or higher is accepted as evidence that a set of items have scaled unidimensionally. AnthroPac reports the CR and the total number of errors and orders the data so that they conform as closely as possible to a perfect Guttman Scale—one that would have a CR of 1.0.

One problem with the CR, however, is that it is sensitive to extreme marginal distributions in terms of both items and individuals so that a high CR can be achieved even in random data (Menzel 1953; McIver and Carmines 1981:50–51). To give an example, suppose an individual randomly answers “yes” to having 90% of the items on a list. Based solely on this fact and without any other knowledge, we can predict with a fair degree of accuracy whether this individual has a “yes” for any given item. This is called *extremeness of individuals*. Similarly, if 90% of respondents answer “yes” to having any one item, predicting that any given individual has this item, in the absence of any other information, is rather easy and can be done within a 10% degree of accuracy. This situation is called *extremeness of items*. In either case, one can predict with a high degree of accuracy the arrangement of data by simply using the category with the largest frequency (i.e., the modal category).

The proportion of correct predictions will be precisely the relative frequency of a given modal category. In other words, although data may have relatively few scale errors—that is, they produce a high CR—they may not necessarily reflect scalability or even departure from randomness. Scalability is such that categories and individuals can be meaningfully arranged from highest to lowest and being able to correctly predict order based solely on marginals undermines any such meaning.

As Menzel (1953) pointed out long ago, the degree of success in reproductions that one attains is a joint result of three things: (1) the extremeness of items, (2) the extremeness of individuals, and (3) the scalability of the items for the given individuals. Therefore, to know if the data exhibit true scalability, we need to control for extreme responses. To deal with this, Menzel (1953) developed the coefficient of scalability (CS), which measures predictability of the scale relative to the level of prediction afforded by consideration solely of the row and column marginals. The formula is the following:  $1 - (\text{number of scale errors} / \text{number of marginal errors})$ .

In this equation, marginal error refers to the number of nonmodal frequencies in a data set.<sup>1</sup> The greater number of marginal errors in proportion to total errors thus increases the CS. Essentially, the CS can be interpreted as a proportion reduction in error (PRE) statistic (McIver and Carmines 1981:50–51). If the number of errors produced by the scale is no fewer than the number you could obtain based on the marginals alone, the CS coefficient is zero. To the extent that the scale has fewer errors than expected by chance, the CS moves toward 1.0. Menzel suggests a CS of .60 or higher as acceptable, but this is not a hard and fast rule.

AnthroPac reports the CR, the CS, and the total number of errors, and orders the data so that they conform as closely as possible to a perfect Guttman Scale—one that would have a CR and CS of 1.0.

### ANALYSIS AND INTEGRATION OF QUALITATIVE DATA

My initial analysis of all twelve items produced a respectable CR of .88 and CS of .38 but, hoping to achieve a CR of .90 and CS of .60, I tried various deletions. There are several ways to do this, but the easiest is to look for the variable with the most errors (Gorden 1977:123; Bernard 1995:296). Using the data management routines in AnthroPac, variables can be deleted in any combination and a new analysis done in seconds. One can continue to organize data to discover the most scalable combination of variables.

Ad hoc data manipulation procedures like this are bound to raise suspicion, but they are very helpful in exploratory research (Goodenough 1968:329; Pelto and Pelto 1978:302), and, in this case, the variable with the most errors is “land” (column 6 in Table 1), with sixty-eight errors. Without this variable, the CR rose to .905 and the CS to .42. I present the ordered list below, with number one being the most commonly held item.

1. mosquito net
2. electricity
3. mattress
4. gas stove
5. television
6. refrigerator
7. stereo
8. canoe
9. outboard motor
10. savings account
11. vehicle

This list of eleven items provides a reasonable scale. However, as every fieldworker knows, numbers alone do not tell the entire story. Some time after collecting these data, I discovered that the concept of land ownership in my research site is not as straightforward as I had originally thought it was. People who hold legal titles to property say that they “own land,” but so do people who are squatters on a piece of land, and many people are squatters in this community. As I had naively measured it, owning land in my field site is not a reliable measure of wealth.

With hindsight, I should not have been surprised to find that the column for land contained the largest number of errors of all the variables in the scale. Squatters who said that they had land were likely to say that they did not have electricity or a television. One household reported having land but not having a mosquito net.

Ethnographic data also provide direction in terms of sorting and manipulating Guttman Scale data. AnthroPac can readily analyze any combination of variables by a simple command, which gives the researcher a large number of possible ways in which to arrange the data. By removing certain items, a higher CR and CS can be achieved. In the present case, I can get a CR of .932 and CS of .505 using six items, .939 and .631 with five items, or .958 and .755 with four. Which and how many variables to use depends on the researcher's question and on the ethnographic context.

Ethnographic work can also lead to insight on things that the numbers may gloss over. Several months after data collection, I found out that many of the local people pirate electricity by simply tapping into a public power line—a very inexpensive process—while others actually pay a monthly fee for electrical service. My survey questions did not account for this distinction, and yet, I was still able to get a CR of .905 and a CS of .42 in the second analysis, which included electricity. Is this a valid variable? I can get slightly higher measures (CR of .910 and CS of .57) by excluding electricity, so should I remove it? These are questions of judgment for the individual researcher. For my final analysis, however, I will remove it since it is not equally accessible to all respondents and would not be a good judge of relative wealth.

Data are never perfect. As Peltó and Peltó (1978:299) observe, "errors in the scale are always to be expected in real human behavior." In my data set, only two households out of the 203 in the sample have vehicles (column 12 in Table 1), and both of those households are scored as having errors of inclusion for vehicles. Should I not include them in the final analysis? And if I do, how should I rank them? Again, one has to return to the ethnographic context for direction. In the case of the two vehicle owners, one is a fish buyer, the other an agricultural merchant. Both of these men (and all eight vehicle owners in the community, for that matter) use their vehicles exclusively for work. I also know that in Ecuador, even the most rundown vehicle is an extremely expensive item, making it cost prohibitive to virtually the entire rural population. So if these men have money to buy and maintain a vehicle, why do they not hold every other variable on the list? A closer look at the data partially reveals what is happening.

It turns out that neither of the two vehicle owners has a boat or outboard motor, items that precede a vehicle on the scale, thus producing errors. When I first ran the survey, I naively thought that a boat was an indispensable asset



in a coastal town, something that everyone aspired to own. Time in the field, however, revealed that this was not the case, and that neither a boat nor a motor are considered necessities. Nor are they always a wise investment. Enlightened with hindsight, I delete both boat and outboard motor from the list and AnthroPac produced a new scale of eight items, with a CR of .93 and CS of .61. This was, of course, an improvement over the previous .910, and was also more consistent with ethnographic reality.

Notwithstanding the increasingly strong CRs and CSs, I am still left with a nagging problem—the relative position of the two vehicle owners. The eight-item scale that I have produced so far places one owner in the second-highest group (having seven items) and the other in the third highest (having six items). The residents of the research site say that only the rich (*los ricos*) have vehicles, so it seems odd that these individuals are not in the uppermost ranking. The question is, what do I do with these “outliers”? Looking at the list of errors, I find that neither has a savings account, and one does not have a stereo.

Perhaps these two people hide their money under the mattress (not such a bad idea considering the recent banking problems in Ecuador) or just do not like music. Should they wind up in a lower wealth rank for these actions? I could ignore the problem altogether since I already have a CR of .932 and a CS of .61, or I can remove the vehicle variable from the list, delete my outliers from the data set, and give them a higher ranking outside of the Guttman Scale (i.e., place them in a group equivalent to having one item more than the highest group in the scale). Given the ethnographic context, I choose to delete vehicle from the scale and rank the two vehicle owners independently, above the highest-ranked group. My final list produces the following scale and a slightly lower CR of .925 and CS of .60.

1. mosquito net
2. mattress
3. gas stove
4. television
5. refrigerator
6. stereo
7. savings account

Deleting outliers may not be a palatable solution to some, especially when it reduces scalability. Indeed, I am reluctant to overrule scale results with a CR well above .90, but it would be equally remiss to ignore the fact that vehicles are a major indicator of wealth in the community.

Alternatively, should I have just ignored the more expensive items altogether when creating the survey? Buying a vehicle in my field site is an impossibility for the majority of households, so a leap into this group is a con-

siderably more difficult, and hence more economically meaningful, achievement than buying a gas stove or a television. For this reason, I have chosen to treat this variable outside of the scale, but ultimately, the nature of the research question determines the import of any one item. In short, both expensive and inexpensive items are important in wealth ranking, but the relative importance of each is contingent on the research goals and the ethnographic context.

This brings to light an important issue—construct validity. Several scholars have questioned the interpretation of Guttman scales, and cautioned against drawing conclusions in the absence of other information (e.g., Graves, Graves, and Kobrin 1968; Marshall and Borthwick 1974). Kronenfeld (1972:259) noted that to be meaningful, any scale requires a “detailed investigation of the underlying variable, the single dimension along which the scaling occurs, and the relation of items (i.e., the scale variables) to it.”

It is not always clear that a scale measures what it is alleged to measure. In the case I am presenting here, the derived scale could be indicative of a growth/dependency process or something else and have nothing to do with wealth. However, looking at the final scale, notice that six of the seven items are local indicators of wealth, as ascertained through many interviews. Some validity is achieved in this way. Also, the items on the scale represent a range of necessity and luxury. And I have cleaned up the scale by removing item dependencies. You need a boat to use an outboard motor and you need a television to use, for example, a VCR. The items in my scale are not related in this way. To be sure, the problem with vehicle ownership casts some doubt on the internal validity of my scale, but the Guttman Scale can be used as one of several techniques to achieve some measurement objective and can be informed by qualitative data to increase credibility.

Scalability does not necessarily denote historical progression in the acquisition of items (Graves, Graves, and Kobrin 1968:319; Kronenfeld 1972). This requires direct observation or questioning of respondents about purchase order. In the absence of these data, I have to rely on my general knowledge of the area. Thus, given the absence of surplus capital in the vast majority of households, it is reasonable to assume that households will purchase goods if and as they have the means to do so (i.e., surplus capital or access to credit) and when it makes sense for them to do so. People in this area also tend to accumulate nonessential items very slowly and so the synchronic picture painted by the Guttman Scale likely does, in fact, reflect the order in which items are accumulated over time.

Finally, confidence in construct validity is increased by testing whether scores from one scale correlate with scores from another scale with which they are expected, on theoretical grounds, to correlate. I have data on the level of education (measured from 0–16 years) for all members of 103 households.

Taking the highest level of education in each of those households, the final seven-item Guttman wealth scale correlates at  $r = .274, p = .01$ . This correlation may indicate that education leads to wealth or vice versa, but either way, we expect a significant correlation between wealth and education. This strengthens the construct validity for the measurement of wealth.

## CONCLUSION

The case I have presented here confirms DeWalt's (1979) finding that it is the smaller household items that tend to form a unidimensional scale. Items that have productive value, and/or that are occupationally specific, may be in a different domain than are the smaller household items. A boat or motor, for example, is not a logical purchase for a household that does not fish.

Logistically, Guttman scaling with AnthroPac is a simple procedure. A yes/no question with regard to the possession of items can be put into any survey and a quick scale of local relative wealth obtained. Respondents find the yes/no format easy and the data are easy for researchers to record and input. The difficult part is the interpretation. This case also illustrates the need for ethnographic data to make decisions about manipulating variables and to interpret results.

## NOTE

1. Nonmodal frequencies refers to the frequency of the least popular answer. For example, in a survey of ten people, if six people answer "yes" to a question, the nonmodal frequency is four. Marginal error is the sum of nonmodal frequencies for all questions.

## REFERENCES

- Adams, A. M, T. G. Evans, R. Mohammed, and J. Farnsworth. 1997. Socioeconomic stratification by wealth ranking: Is it valid? *World Development* 25 (7): 1165–72.
- Bernard, H. R. 1995. *Research methods in anthropology: Qualitative and quantitative approaches*. 2d ed. London: AltaMira.
- Borgatti, S. 1992. *AnthroPac methods guide*. Columbia, SC: Analytic Technologies.
- Carneiro, R. L. 1962. Scale analysis as an instrument for the study of cultural evolution. *Southwestern Journal of Anthropology* 18:149–69.
- . 1973. Scale analysis, evolutionary sequences, and the rating of cultures. In *A handbook of method in cultural anthropology*, edited by R. Naroll and R. Cohen, 834–71. New York: Columbia University Press.

- DeWalt, B. R. 1979. *Modernization in a Mexican ejido*. New York: Cambridge University Press.
- Edwards, A. 1948. On Guttman's scale analysis. *Educational and Psychological Measurement* 8:313–18.
- Goodenough, W. H. 1944. A technique for scale analysis. *Educational and Psychological Measurement* 4:179–90.
- . 1968. Comments on Graves et al. *Current Anthropology* 10:329.
- Gorden, R. 1977. *Unidimensional scaling of social variables: Concepts and procedures*. New York: Free Press.
- Grandin, B. E. 1988. *Wealth ranking in smallholder communities: A field manual*. Nottingham, UK: Russell Press, Ltd.
- Graves, T. D., N. B. Graves, and M. F. Kobrin. 1968. Historical inferences from Guttman scales: The return of age-area magic. *Current Anthropology* 10:317–27, 334–38.
- Guttman, L. 1944. A basis for scaling qualitative data. *American Sociological Review* 9:139–50.
- . 1950. The basis for scalogram analysis. In *Measurement of prediction*, edited by S. Stouffer et al., 60–90. Princeton, NJ: Princeton University Press.
- Hammel, E. A. 1962. Social rank and evolutionary position in a coastal Peruvian village. *Southwestern Journal of Anthropology* 18:199–215.
- Johnson, A. 1998. Guttman scaling: An analysis of Matsigenki men's manufacturing skills. In *Using methods in the field: A practical introduction and casebook*, edited by V. de Munck and E. Sobo, 111–20. Walnut Creek, CA: AltaMira.
- Kay, P. 1964. A Guttman scale model of Tahitian consumer behavior. *Southwestern Journal of Anthropology* 20:160–67.
- Kronenfeld, D. B. 1972. Guttman scaling: Problems of conceptual domain, unidimensionality, and historical inference. *Man* 7:255–76.
- Marshall, M., and M. Borthwick. 1974. Consensus, dissensus and Guttman scales: The Namoluk case. *Journal of Anthropological Research* 30:252–70.
- McIver, J. P., and E. G. Carmines. 1981. *Unidimensional scaling*. Beverly Hills, CA: Sage.
- Menzel, H. 1953. A new coefficient for scalogram analysis. *Public Opinion Quarterly* 17:268–80.
- Pelto, P. J. 1973. *The snowmobile revolution: Technology and social change in the Arctic*. Menlo Park, CA: Cummings.
- Pelto, P. J., and G. H. Pelto. 1978. *Anthropological research: The structure of inquiry*. Cambridge, UK: Cambridge University Press.
- Werner, D. 1985. Psycho-social stress and the construction of a flood-control dam in Santa Catarina, Brazil. *Human Organization* 44:161–66.

*GREG GUEST received his Ph.D. in anthropology from the University of Georgia, Athens. He is currently employed in the experience modeling department of Sapient Corporation in San Francisco, California, where he helps develop and employ anthropological methods to better understand human/technology interfacing.*