Muehrcke P C, Muehrcke J O 1992 *Map Use: Reading, Analysis, Interpretation*, 3rd edn. JP Publications, Madison, WI
Openshaw S 1983 *The Modifiable Areal Unit Problem*. Geo Books, Norfolk, UK

D. R. Montello

Scaling and Classification in Social Measurement

Social measurements translate observed characteristics of individuals, events, relationships, organizations, societies, etc. into symbolic classifications that enable reasoning of a verbal, logical, or mathematical nature. Qualitative research and censuses together define one realm of measurement, concerned with assignment of entities to classification categories embedded within taxonomies and typologies. Another topic in measurement involves scaling discrete items of information such as answers to questions so as to produce quantitative measurements for mathematical analyses. A third issue is the linkage between social measurements and social theories.

1. *Classifications*

Classification assimilates perceived phenomena into symbolically labeled categories. Anthropological studies of folk classification systems (D'Andrade 1995) have advanced understanding of scientific classification systems, though scientific usages involve criteria that folk systems may not meet entirely.

Two areas of social science employ classification systems centrally. Qualitative analyses such as ethnographies, histories, case studies, etc. offer classifications—sometimes newly invented—for translating experiences in unfamiliar cultures or minds into familiar terms. Censuses of individuals, of occurrences, or of aggregate social units apply classifications—usually traditional—in order to count entities and their variations. Both types of work depend on theoretical constructions that link classification categories.

1.1 Taxonomies

Every classification category is located within a taxonomy. Some more general categorization, Y, determines which entities are in the domain for the focal categorization, X; so an X always must be a kind of Y. 'X is a kind of Y' is the linguistic frame for specifying taxonomies. Concepts constituting a taxonomy form a logic tree, with subordinate elements implying superordinate items.

Taxonomic enclosure of a classification category is a social construction that may have both theoretical and practical consequences. For example, if only violent crimes are subject to classification as homicides, then 'homicide' is a kind of 'violent crime,' and deaths caused by executive directives to release deadly pollution could not be homicides.

1.2 Typologies

A typology differentiates entities at a particular level of a taxonomy in terms of one or more of their properties. The differentiating property (sometimes called a feature or attribute) essentially acts as a modifier of entities at that taxonomic level. For example, in the USA kinship system siblings are distinguished in terms of whether they are male or female; in Japan by comparison, siblings are schematized in terms of whether they are older as well as whether they are male or female.

A scientific typology differentiates entities into types that are exclusive and exhaustive: every entity at the relevant taxonomic level is of one defined type only, and every entity is of some defined type. A division into two types is a dichotomy, into three types a trichotomy, and into more than three types a polytomy.

Polytomous typologies are often constructed by crossing multiple properties, forming a table in which each cell is a theoretical type. (The crossed properties might be referred to as variables, dimensions, or factors in the typology.) For example, members of a multiplex society have been characterized according to whether they do or do not accept the society's goals on the one hand, and whether they do or do not accept the society's means of achieving goals on the other hand; then crossing acceptance of goals and means produces a fourfold table defining conformists and three types of deviants.

Etic–emic analysis involves defining a typology with properties of scientific interest (the etic system) and then discovering ethnographically which types and combinations of types are recognized in folk meanings (the emic system). Latent structure analysis statistically processes observed properties of a sample of entities in order to confirm the existence of hypothesized types and to define the types operationally.

1.3 Aggregate Entities

Aggregate social entities such as organizations, communities, and cultures may be studied as unique cases, where measurements identify and order internal characteristics of the entity rather than relate one aggregate entity to another.

A seeming enigma in social measurement is how aggregate social entities can be described satisfactorily on the basis of the reports of relatively few informants, even though statistical theory calls for substantial samples of respondents to survey populations. The key is that informants all report on the same thing—a single culture, community, or organization—whereas respondents in a social survey typically report on diverse things—their own personal characteristics, beliefs, or experiences. Thus, reports from informants serve as multiple indicators of a single state, and the number needed depends on how many observations are needed to define a point reliably, rather than how many respondents are needed to describe a population's diversity reliably. As few as seven expert informants can yield reliable descriptions of aggregate social entities, though more are needed as informants' expertise declines (Romney et al. 1986). Informant expertise correlates with greater intelligence and experience (D'Andrade 1995) and with having a high level of social integration (Thomas and Heise 1995).

2. *Relations*

Case grammar in linguistics defines events and relationships in terms of an actor, action, object, and perhaps instrumentation, setting, products, and other factors as well. Mapping sentences (Shye et al. 1994) apply the case grammar idea with relatively small lists of entities in order to classify relational phenomena within social aggregates. For example, interpersonal relations in a group can be specified by pairing group members with actions such as loves, admires, annoys, befriends, and angers. Mapping sentences defining the relations between individuals or among social organizations constitute the measurement model for social network research.

3. *Scaling*

Quantitative measurements differentiate entities at a given taxonomic level—serving like typological classifications, but obtaining greater logical and mathematical power by ordering the classification categories.

An influential conceptualization (Stevens 1951) posited four levels of quantification in terms of how numbers relate to classification categories. *Nominal* numeration involves assigning numbers arbitrarily simply to give categories unique names, such as 'batch 243.' An *ordinal* scale's categories are ordered monotonically in terms of greater-than and less-than, and numbering corresponds to the rank of each category. Numerical ranking of an individual's preferences for different foods is an example of ordinal measurement. Differences between categories can be compared in an *interval* scale, and numbers applied to categories reflect degrees of differences. Calendar dates are an example of interval measurements—we know from their birth years that William Shakespeare was closer in time to Geoffrey Chaucer than Albert Einstein was to Isaac Newton. In a *ratio* scale categories have magnitudes that are whole or fractional multiples of one another, and numbers assigned to the categories represent these magnitudes. Population sizes are an example of ratio measurements—knowing the populations of both nations, we can say that Japan is at least 35 times bigger than Jamaica.

A key methodological concern in psychometrics (Hopkins 1998) has been: how do you measure entities on an interval scale given merely nominal or ordinal information?

3.1 *Scaling Dichotomous Items*

Nominal data are often dichotomous yes–no answers to questions, a judge's presence–absence judgments about the features of entities, an expert's claims about truth–falsity of propositions, etc. Answers of *yes*, *present*, *true*, etc. typically are coded as 'one' and *no*, *absent*, *false*, etc. as 'zero.' The goal, then, is to translate zero-one answers for each case in a sample of entities into a number representing the case's position on an interval scale of measurement.

The first step requires identifying how items relate to the interval scale of measurement in terms of a graph of the items' characteristic curves. The horizontal axis of such a graph is the interval scale of measurement, confined to the practical range of variation of entities actually being observed. The vertical axis indicates probability that a specific dichotomous item has the value one for an entity with a given position on the interval scale of measurement. An item's characteristic curve traces the changing probability of the item having the value one as an entity moves from having a minimal value on the interval scale to having the maximal value on the interval scale.

Item characteristic curves have essentially three different shapes, corresponding to three different formulations about how items combine into a scale.

Spanning items have characteristic curves that essentially are straight lines stretching across the range of entity variation. A spanning item's line may start as a low probability value and rise to a high probability value, or fall from a high value to a low value. A rising line means that the item is unlikely to have a value of one with entities having a low score on the interval scale; the item is likely to have a score of one for entities having a high score on the interval scale; and the probability of the item being valued at one increases regularly for entities between the low and high positions on the scale.

Knowing an entity's value on any one spanning item does not permit assessing the entity's position along the interval scale. However, knowing the entity's values on multiple spanning items does allow an estimate of positioning to be made. Suppose heuristically that we are working with a large number of equivalent spanning items, each having an item characteristic curve that starts at probability 0.00 at the minimal point of the interval scale, and rises in a straight line to probability 1.00 at the maximal point of the interval scale. The probability of an item being valued at one can be estimated from the observed proportion of all these items that are valued at one—which is simply the mean item score when items are scored zero-one. Then we can use the characteristic curve for the items to find the point on the interval scale where the entity must be positioned in order to have the estimated item probability. This is the basic scheme involved in construction of *composite scales*, where an averaged or summated score on multiple items is used to estimate an entity's interval-scale value on a dimension of interest (Lord and Novick 1968).

The more items that are averaged, the better the estimate of an entity's position on the interval scale. The upper bound on number of items is pragmatic, determined by how much precision is needed and how much it costs to collect data with more items. The quality of the estimate also depends on how ideal the items are in terms of having straight-line characteristic curves terminating at the extremes of probability. Irrelevant items with a flat characteristic curve would not yield an estimate of scale position no matter how many of them are averaged, because a flat curve means that the probability of the item having a value of one is uncorrelated with the entity's position on the interval scale. Inferences are possible with scales that include relevant but imperfect items, but more items are required to achieve a given level of precision, and greater weight needs to be given to the more perfect items.

Declivitous items have characteristic curves that rise sharply at a particular point on the horizontal axis. Idealized, the probability of the item having a value of one increases from 0.00 to the left of the inflection point to 1.00 to the right of the inflection point; or alternatively the probability declines from 1.00 to 0.00 in passing the inflection point. Realistically, the characteristic curve of a declivitous item is S-shaped with a steep rise in the middle and graduated approaches to 0.00 at the bottom and to 1.00 at the top.

The value of a single declivitous item tells little about an entity's position along the interval scale. However, an inference about an entity's scale position can be made from a set of declivitous items with different inflection points, or difficulties, that form a *cumulative scale*. Suppose heuristically that each item increases stepwise at its inflection point. Then for an entity midway along the horizontal axis, items at the left end of the scale will all have the value of one, items at the right end of the scale will all have the value zero, and the entity's value on the interval scale is between the items with a score of one and the items with a score of zero.

If the items' inflection points are evenly distributed along the interval scale, then the sum of items' zero-one scores for an entity constitutes an estimate of where the entity is positioned along the interval scale. That is, few of the items have a value of one if the entity is on the lower end of the interval scale, and many of the items are valued at one if the entity is at the upper end of the interval scale. This is the basic scheme involved in Guttman scalogram analysis (e.g., see Shye 1978, Part 5). On the other hand, we might use empirical data to estimate the position of each item's inflection point on the interval scale, while simultaneously estimating entity scores that take account of the item difficulties. This is the basic scheme involved in scaling with Rasch models (e.g., Andrich 1988).

Entities' positions on the interval scale can be pinned down as closely as desired through the use of more declivitous items with inflection points spaced closer and closer together. However, adding items to achieve more measurement precision at the low end of the interval scale does not help at the middle or the high end of the interval scale. Thus, obtaining high precision over the entire range of variation requires a large number of items, and it could be costly to obtain so much data. Alternatively, one can seek items whose characteristic curves rise gradually over a range of the interval scale such that sequential items on the scale have overlapping characteristic curves, whereby an entity's position along the interval scale is indicated by several items.

Regional items have characteristic curves that rise and fall within a limited range of the interval scale. That is, moving an entity up the interval scale increases the probability of a particular item having a value of one for a while, and then decreases the probability after the entity passes the characteristic curve's maximum value. For example, in a scale measuring prejudice toward a particular ethnic group, the probability of agreeing with the item 'they require equal but separate facilities' increases as a person moves away from an apartheid position, and then decreases as the person moves further up the scale toward a nondiscriminatory position. A regional item's characteristic curve is approximately bell-shaped if its maximum is at the middle of the interval scale, but characteristic curves at the ends of the scale are subject to floor and ceiling clipping, making them look like declivitous items.

If an entity has a value of one on a regional item, then the entity's position along the interval scale is known approximately, since the entity must be positioned in the part of the scale where that item has a nonzero probability of being valued at one. However, a value of zero on the same item can result from a variety of positions along the interval scale and reveals little about the entity's position. Thus, multiple regional items have to be used to assess positions along the whole range of the scale. The items have to be sequenced relatively closely along the scale with overlapping characteristic curves so that no entity will end up in the noninformative state of having a zero value on all items.

One could ask judges to rate the scale position of each item, and average across judges to get an item score; then, later, respondents can be scored with the average of the items they accept. This simplistic approach to regional items was employed in some early attempts to measure social attitudes. Another approach is statistical unfolding of respondents' choices of items on either side of their own positions on the interval scale in order to estimate scale values for items and respondents simultaneously (Coombs 1964).

Item analysis is a routine aspect of scale construction with spanning items, declivitous items, or regional items. One typically starts with a notion of what one wants to measure, assembles items that should relate to the dimension, and tests the items in order to select the items that work best. Since a criterion measurement that can be used for assessing item quality typically is lacking, the items as a group are assumed to measure what they are supposed to measure, and scores based on this assumption are used to evaluate individual items.

Items in a scale are presumed to measure a single dimension rather than multiple dimensions. Examining the dimensionality assumption brings in additional technology, such as component analysis or factor analysis in the case of spanning items, multidimensional scalogram analysis in the case of declivitous items, and nonmetric multidimensional scaling in the case of regional items. These statistical methods help in refining the conception of the focal dimension and in selecting the best items for measuring that dimension.

3.2 Ordered Assessments

Ordinal data—the starting point for a three-volume mathematical treatise on measurement theory (Krantz et al. 1971, Suppes et al. 1989, Luce et al. 1990)—may arise from individuals' preferences, gradings of agreement with statements, reckonings of similarity between stimuli, etc. Conjoint analysis (Luce and Tukey 1964, Michell 1990) offers a general mathematical model for analyzing such data. According to the conjoint theory of measurement, positions on any viable quantitative dimension are predictable from positions on two other quantitative dimensions, and this assumption leads to tests of a dimension's usefulness given just information of an ordinal nature. For example, societies might be ranked in terms of their socioeconomic development and also arrayed in terms of the extents of their patrifocal technologies (like herding) and matrifocal technologies (like horticulture), each of which contributes additively to socioeconomic development. Conjoint analyses could be conducted to test the meaningfulness of these dimensions, preliminarily to developing interval scales of socioeconomic development and patrifocal and matrifocal technologies. Specific scaling methodologies, like Rasch scaling and nonmetric multidimensional scaling, can be interpreted within the conjoint analysis framework.

Magnitude estimations involve comparisons to an anchor, for example: 'Here is a reference sound.… How loud is this next sound relative to the reference sound.' Trained judges using such a procedure can assess intensities of sensations and of a variety of social opinions on ratio scales (Stevens 1951, Lodge 1981). Comparing magnitude estimation in social surveys to the more common procedure of obtaining ratings on category scales with a fixed number of options, Lodge (1981) found that magnitude estimations are more costly but more accurate, especially in registering extreme positions.

Rating scales with bipolar adjective anchors like good–bad are often used to assess affective meanings of perceptions, individuals, events, etc. Such scales traditionally provided seven or nine answer positions between the opposing poles with the middle position defined as neutral. Computerized presentations of such scales with hundreds of rating positions along a graphic line yield greater precision by incorporating some aspects of magnitude estimation. Cross-cultural and cross-linguistic research in dozens of societies has demonstrated that bipolar rating scales align with three dimensions—evaluation, potency, and activity (Osgood et al. 1975). An implication is that research employing bipolar rating scales should include scales measuring the standard three dimensions in order to identify contributions of these dimensions to rating variance on other bipolar scales.

4. Measurements and Theory

Theories and measurements are bound inextricably. In the first place, taxonomies and typologies—which are theoretical constructions, even when rooted in folk classification systems—are entailed in defining which entities are to be measured, so are part of all measurements.

Second, scientists routinely assume that any particular measurement is wrong to some degree—even a measurement based on a scale, and that combining multiple measurements improves measurement precision. The underlying premise is that a true value exists for that which is being measured, as opposed to observed values, and that theories apply to true values, not ephemeral observed values. In essence, theories set expectations about what should be observed, in contrast to what is observed, and deviation from theoretical expectations is interpreted as measurement error (Kyburg 1992). A notion of measurement error as deviation from theoretical expectations is widely applicable, even in qualitative research (McCullagh and Behan 1984, Heise 1989).

Third, the conjoint theory of measurement underlying many current measurement technologies requires theoretical specification of relations between variables before the variables' viability can be tested or meaningful scales constructed. This approach is in creative tension with traditional deductive science, wherein variable measurements are gathered in order to determine if relations among variables exist and theories are correct. In fact, a frequent theme in social science during the late twentieth century was that measurement technology had to improve in order to foster the growth of more powerful theories. It is clear now that the dependence between measurements and theories is more bidirectional than was supposed before the development of conjoint theory.

See also: Classification: Conceptions in the Social Sciences; Dimensionality of Tests: Methodology; Factor Analysis and Latent Structure: IRT and Rasch Models; Test Theory: Applied Probabilistic Measurement Structures

Bibliography

Andrich D 1988 *Rasch Models for Measurement*. Sage, Newbury Park, CA

Coombs C H 1964 *A Theory of Data*. John Wiley & Sons, New York

D'Andrade R 1995 *The Development of Cognitive Anthropology*. Cambridge University Press, New York

Heise D R 1989 Modeling event structures. *Journal of Mathematical Sociology* **14**: 139–69

Hopkins K D 1998 *Educational and Psychological Measurement and Evaluation*, 8th edn. Allyn and Bacon, Boston

Krantz D H, Luce R D, Suppes P, Tversky A 1971 *Foundations of Measurement. Volume 1: Additive and Polynomial Representations*. Academic Press, New York

Kyburg Jr. H E 1992 Measuring errors of measurement. In: Savage C W, Ehrlich P (eds.) *Philosophical and Foundational Issues in Measurement*. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 75–91

Lodge M 1981 *Magnitude Scaling: Quantitative Measurement of Opinions*. Sage, Beverly Hills, CA

Lord F N, Novick M R 1968 *Statistical Theories of Mental Test Scores*. Addison-Wesley, Reading, MA

Luce R D, Krantz D H, Suppes P, Tversky A 1990 *Foundations of Measurement. Volume 3: Representation, Axiomatization, and Invariance*. Academic Press, New York

Luce R D, Tukey J W 1964 Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology* **1**: 1–27

McCullagh B C, Behan C 1984 *Justifying Historical Descriptions*. Cambridge University Press, New York

Michell J 1990 *An Introduction to the Logic of Psychological Measurement*. Lawrence Erlbaum Associates, Hillsdale, NJ

Osgood C, May W H, Miron M S 1975 *Cross-cultural Universals of Affective Meaning*. University of Illinois Press, Urbana, IL

Romney A K, Weller S C, Batchelder W H 1986 Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist* **88**: 313–38

Shye S (ed.) 1978 *Theory Construction and Data Analysis in the Behavioral Sciences*, 1st edn. Jossey-Bass, San Francisco

Shye S, Elizur D, Hoffman M 1994 *Introduction to Facet Theory: Content Design and Intrinsic Data Analysis in Behavioral Research*. Sage, Thousand Oaks, CA

Stevens S S 1951 Mathematics, measurement, and psychophysics. In: Stevens S S (ed.) *Handbook of Experimental Psychology*. John Wiley and Sons, New York, pp. 1–49

Suppes P, Krantz D H, Luce R D, Tversky A 1989 *Foundations of Measurement. Volume 2: Geometrical, Threshold, and Probabilistic Representations*. Academic Press, New York

Thomas L, Heise D R 1995 Mining error variance and hitting pay-dirt: Discovering systematic variation in social sentiments. *The Sociological Quarterly* **36**: 425–39

D. R. Heise

Scaling: Correspondence Analysis

In the early 1960s a dedicated group of French social scientists, led by the extraordinary scientist and philosopher Jean-Paul Benzécri, developed methods for structuring and interpreting large sets of complex data. This group's method of choice was correspondence analysis, a method for transforming a rectangular table of data, usually counts, into a visual map which displays rows and columns of the table with respect to continuous underlying dimensions. This article introduces this approach to scaling, gives an illustration and indicates its wide applicability. Attention is limited here to the descriptive and exploratory uses of correspondence analysis methodology. More formal statistical tools have recently been developed and are described in *Multivariate Analysis: Discrete Variables (Correspondence Models)*.

1. *Historical Background*

Benzécri's contribution to data analysis in general and to correspondence analysis in particular was not so much in the mathematical theory underlying the methodology as in the strong attention paid to the graphical interpretation of the results and in the broad applicability of the methods to problems in many contexts. His initial interest was in analyzing large sparse matrices of word counts in linguistics, but he soon realized the power of the method in fields as diverse as biology, archeology, physics, and music. The fact that his approach paid so much attention to the visualization of data, to be interpreted with a degree of ingenuity and insight into the substantive problem, fitted perfectly the *esprit géometrique* of the French and their tradition of visual abstraction and creativity.

Originally working in Rennes in western France, this group consolidated in Paris in the 1970s to become an influential and controversial movement in post-1968 France. In 1973 they published the two fundamental volumes of, *L'Analyse des Données* (Data Analysis), the first on *La Classification*, that is,

International Encyclopedia of the Social & Behavioral Sciences ISBN: 0-08-043076-7