Shye, S. (2014). Faceted Smallest Space Analysis (Faceted SSA; FSSA). In: Michalos AC (Ed.). *Encyclopedia of Quality of Life and Well-Being Research.* Dordrecht, Netherlands: Springer, pp 2129-2134.

Faceted Smallest Space Analysis (Faceted SSA; FSSA)

Synonyms

FSSA, Faceted MDS, Faceted Multidimensional Scaling, Confirmatory SSA, MDS, Multidimensional Scaling, Non-metric MDS, Geometric Representation, Spatial Representation, Similarity Analysis, Perceptual Mapping, Cognitive Maps, Structural Analysis, FSSAWIN, Regional hypotheses.

Definition

FSSA (Faceted SSA or Faceted MDS) is a multivariate data analytic procedure for inferring the structure of a content-universe under study (containing a large, possibly infinite number of variables) from the correlation matrix of a representative sample of observed variables taken from that content-universe. FSSA is useful for constructing theories and testing hypotheses in research domains, such as Quality of Life, characterized by a multitude of variables not all of which may be observed. FSSA consists of

(1) Processing data by some non-metric similarity analysis technique such as non-metric MDS (Kruskal, 1964; Borg and Groenen, 2005) or SSA-I (Guttman, 1968; Lingoes, 1973), yielding a geometric space where observed variables are mapped as points with the pairwise inter-point distances (inversely) ranked by the correlations between the corresponding pairs of variables; and then
(2) Finding a one-to-one correspondence between a *content-classification* of the variables (a content-*facet*) and a *partition into regions* (or hyper-regions) of the resultant geometric space, so that every class of variables is located in one region of that partition (Shye, 1971; 1978a; 1998; 1999; Shye & Elizur, 1994).

More generally, the objects mapped into the geometric space need not be variables but may be any other objects of a given type, between every pair of which a similarity measure – or a dissimilarity measure – can be defined. FSSA has been found useful in a variety of QOL and QOL-related studies (e.g., see references in Systemic Quality of Life model (SQOL).

Description

The Analysis of Multivariate Data by MDS: Background

Multidimensional scaling (MDS) is a family of statistical procedures used for analyzing and visually representing multivariate data. Given a set of objects and a matrix of similarity or dissimilarity measures between each of pair of the objects in the set, an MDS procedure aims to represent the objects as points in a geometric space of a given (usually small) dimensionality as well as possible, such that the greater the dissimilarity between two objects, the greater the distance between them. MDS may be metric or non-metric. In the latter case, MDS seeks distances whose rank-order matches that of the corresponding dissimilarities. The mapped objects are often observed variables whose (zero-order) correlation coefficients are taken as similarity measures between pairs of variables. Different algorithms have been created for MDS, based on different procedures and loss functions (Borg and Groenen, 2005). One of the popular MDS procedures is Guttman's (1968) SSA.

Typically, MDS is used for visualizing data – e.g. the set of all associations among research variables– for data mining, and for relating content similarity of observed variables to their geometric proximity (see, e.g., Foa's (1965) contiguity principle). Emphasis is on accuracy of presentation (low values of the loss function employed, such as Kruskal's (1964) *stress*, or Guttman's (1968) *Coefficient of Alienation*) and in the interpretation stage, attention is often given to clusters of the objects in space.

FSSA, an offshoot of MDS/SSA, differs from traditional MDS/SSA in the interpretative stage and offers a novel computational procedure for partitioning the resultant space by content criteria. FSSA is a data analytic procedure based on a paradigm of semantic continuity: It is derived from a conception that pictures statistical data in behavioral and social research as *continuous* spaces. Hence, FSSA relies on research design, and suggests kinds of hypotheses (and provides tools for their testing) that ensue from that conception.

Semantic Continuity in Behavioral Data

In the statistical study of QOL, as in many domains of behavioral and social research, the number of variables in is very large, even infinite. Thus, for example, in principle, the number of questionnaire items of well-being is potentially endless (considering that any variation in the phrasing of a question defines a separate item, or variable); and so is the number of test items in a cognitive ability or intelligence testing. This observation, together with the idea of mapping variables into a geometric space, as effected by the MDS family of statistical procedure, prompted the formulation of the following:

Continuity Principle: The mapping of variables from the content universe to a topological manifold is one-to-one and ‘onto’. (Shye, 1971; 1998; 1999; Shye & Elizur, 1994).

Specifically, the content universe of a studied domain is pictured as a continuum. Every point in the topological manifold representing that universe represents a possible variable of that domain. Moreover, the mapping of variables into a suitable topological manifold is continuous and has a continuous inverse mapping (i.e., the mapping is 'homeomorphic’).

The Continuity Principle (Shye, 1971; 1998; 1999; Shye & Elizur, 1994) intuitively means that continuity in meanings of a system of variables is reflected by continuity in the geometric space, and vice versa. And hence, that the actually *observed* variables in a given study are no more than a *sample* from the infinite universe of variables, as indeed they typically are in behavioral and psychometric research. The continuity principle has led to a new *scientific imagery* within the framework of modern facet theory (Shye, 1998; 1999; Shye & Elizur, 1994). This scientific imagery differs greatly from traditional approaches that tend to treat the observed variables as important in and by themselves (rather than as a sample that represents an unobservable whole); and it suggests and even determines new kinds of questions and hypotheses that observed behavioral sciences could fruitfully deal with.

In practice, only a finite number of variables can be observed. This calls for techniques for *sampling variables* from the universe of the investigated contents, and for *making inferences* from that sample to the entire universe. This is much in analogy with the statistical sampling and inferential techniques regarding the investigated population.

Faceted Sampling of Observations

Common-Meaning Range. A prior condition for analyzing data by FSSA is that all variables included in the analysis have a Common-Meaning-Range (CMR). This means that numerically coded variables such as answers to psychological tests, responses to attitude questionnaire items, or QOL assessments be all ordered from high to low in the same direction according to the concept under study (e.g., well-being, cognitive ability, favorable attitude towards an object. But the number of possible response categories need not be the same.) This ensures that the correlations between variables are computed with unequivocal sign and that, in line with scientific imagery described above, the space obtained, uniformly represents the studied concept (i.e. *no* part of that space represents the *opposite* of the concept).

Sampling of variables: **faceted design.** It is usually impossible to conduct a probability sampling of variables (as is often done when sampling respondents). Hence sampling of variables by facet design resorts a method akin to *stratified sampling*: the conceived

content universe, namely, the set of all conceivable variables pertaining to the studied concept, is classified by the contents of the variables. A classification of variables by a given content criterion is called a *content-facet*. The choice of a content-facet reflects the expertise, insights and scope of acquaintance of the researcher with the studied domain (Shye, 1999; Shye and Elizur, 1994). Then, attempt is made to have all classes of variables well represented in the sample of variables. And, if more than one content-facet is desired, an attempt is made to have every cell of the cross-classification represented by variables to be observed.

Faceted Inference of Content Structure

Identifying regions in space. A sample of observed variables representative of a well defined domain – having a Common Meaning Range and created by faceted design as described above – can be meaningfully processed by one of the MDS/SSA techniques for the purpose of making inferences from the sample of the observed variables to the entire content universe that they represent. The obtained geometric space is regarded as a topological manifold that may be partitioned into regions that correspond to the classes of a facet defined in advance in the faceted design; or by a newly conceived facet suggested by the distribution of the variables in MDS- space. More than one facet may be so examined, with each facet corresponding to a different partition pattern of the space. When one or more facets are found to correspond to clear partitions, it becomes meaningful to talk about the *'structure'* of a concept such as intelligence or quality of life (see, e.g., Systemic Quality of Life model (SQOL).

The meaning of 'structure'. The term 'structure' in the present context has a well-defined meaning (Shye, 1978b p. 338), comprising the following aspects (i) the optimal MDS *dimensionality* required for the concept representation; (ii) the concept *components* validated by regions (hyper-regions) in the MDS space; (iii) *relationships* among the concept components, where these relationships are depicted and cast in geometric terms, i.e., in terms of the relative spatial orientation of the respective validated regions (the validated partition patterns).

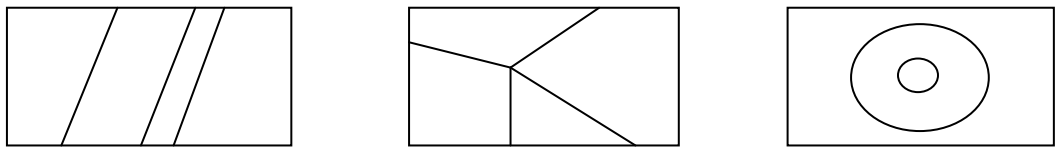
Types of partition patterns. In a given dimensionality, the number of possible partition patterns is of course limitless. Here we describe three canonical partition patterns that may occur in a 2-dimensional space which, because of their simplicity, (i) can be easily generalized to higher dimensionalities, (ii) can be easily interpreted in terms of the research substance, and (iii) can be expected to recur in replications.

The axial partition pattern. A partition into simply ordered stripes by parallel lines (Figure 1a).

The angular partition pattern. A partition into circularly ordered sectors by radii radiating from a common center (Figure 1b).

The radial partition pattern. A partition into simply ordered concentric rings by concentric circles having a common center (Figure 1c).

Figure 1. Canonical Partition Patterns in 2-dimensional Space



a. Axial **b. Angular** **c. Radial**

Partitioning the MDS space by a content-classification of the variables, though a qualitative feature of the data, constitutes in effect a new kind of statistic whose 'values' are specific partition patterns. A partition pattern may be hypothesized in advance (confirmatory FSSA) or detected *ex-factum* (exploratory FSSA).

Confirmatory FSSA. While partition patterns can often be detected or verified by inspection and drawn by hand, an algorithm testing for each of the canonical patterns has been created and programmed (Kingsley, 1991; Shye, 1991; Borg and Shye, 1995). Given a set of variables, pre-classified by the researcher, and given the 2-d MDS/SSA of this set of variables, FSSA finds, for each of the three canonical partition patterns types, the one specific partition of that type that best separates the points in the space obtained; so that points pertaining to a class of variables fall as closely as possible into one region. The goodness-of-fit of the obtained separation is called the Separation Index (SI) and is computed by:

$$SI = 1 - (\text{loss function}) / (\text{normalizing function})$$

where the *loss function* is made up of the sums of the distances of each deviant point from its prescribed region, and where a deviant point is one which does not fall in the region assigned to its class.

The *normalizing function* represents the typical loss function for a set of points randomly (uniformly) distributed in the square 100 * 100. The normalized loss function (i.e., the ratio) falls roughly on a scale between 0 and 1, where 0 represents a perfect separation, with each class of points falling entirely into its assigned region. Hence, the resulting Separation Index falls between 0 and 1, with 1 indicating a perfect separation by the pre-specified content classification of the variables. (Note that the loss function and the Separation Index are not based on the *number* of deviations but on their *sizes*.)

An intuitive interpretation of the value of the Separation Index, say, SI=0.95, could be this: The sum total of the deviations is (about) 5% of what they would have been if the points were scattered at random.

It is important to realize that with this procedure, FSSA produces *two* measures of goodness-of-fit. The one (represented by *low* values in one of the loss functions of MDS/SSA, e.g., *Stress*, or *coefficient-of-alienation*), assesses how well the distances in the obtained MDS/SSA-space of the given dimensionality reflect the input similarities (e.g., correlations). The other, (represented by *high* values of SI) assesses how well the obtained space-partition separates variables according to their input content classification (content facet).

Running FSSA. In contrast with other statistical procedures, FSSA favors a multitude of variables. In fact, a sufficiently large number of variables is a *necessary* condition for running FSSA meaningfully. (Processing 20-90 variables is not uncommon.) But it is not a *sufficient* condition: It is also necessary that the variables be sufficiently well spread in the resultant space so that partitions may be inferred unambiguously. This is more likely the better is the sampling of the variables (see above).

Unsatisfactory (i.e., high) stress/coefficient-of-alienation in MDS/SSA of a given dimensionality may mean either that the content universe represented by the variables requires a higher dimensionality (alternatively one may focus on a sub-content universe), or that there is much noise (content- fluctuations) present. In the latter case, remaining in the lower dimensionality serves the useful purpose of ironing out perturbations, which facilitates concentrating on the essential structure of the data (Shye, 1985 p.164). SI, on the other hand, serves to assess the validity of the structure and more generally of the substantive hypothesis or theory that the researcher wishes to test. If a regional hypothesis is confirmed across replications, it can have a compelling power and scientific significance, regardless of possible inaccuracies of the MDS correlation representation, which may reflect noise. (Details on how to run FSSA, and in particular how to decide on the suitable dimensionality, can be found in Shye and Elizur (1994).)

FSSA compared with Factor Analysis (FA). The two techniques, Faceted SSA and Factor Analysis, have been compared by Guttman (1982) and by Shye (1988). An important difference between FSSA and FA is this: While FA seeks to structure the *observed* variables and does so by identifying ‘factors’; FSSA seeks to infer the structure of the *entire* content universe, including unobserved variables, and does so by identifying regions, using the observed variables as but a *sample* that provides clues to the content universe structure. This difference can be dramatic: For example, a cluster of variables that would define a factor in FA, may be split into two different regions by FSSA. And conversely, two variables that are far apart and pertain to two different factors in FA, may pertain to one and the same region in FSSA, even if the space between them is empty, a fact which FSSA would dismiss as an artifact of the procedure by which variables had been sampled.

External variables in FSSA. Researchers often wish to explore the relationship between the investigated content universe and an 'external variable' – one that does not (or not necessarily) belong to that universe. For this purpose, Denesh & Shye (1993) have proposed a simple procedure that is most in line with the spirit of the *Facet Approach*: On the FSSA interpreted map, one tries to identify a region or regions where the correlations of the 'internal' variables (those pertaining to the investigated content universe and have been analyzed by FSSA) are relatively high (above some specified value). Then, an attempt is made to interpret that region in substantive terms as a sub-content universe of the originally investigated content universe.

An instructive example of faceted external variable analysis that served to identify the most important Quality of Life components for *successful immigration* (as external variable) can be found in Benish-Weismen and Shye (2011). In that study, the high correlation region suggested a compact QOL-sub-universe coined in the language of the Systemic Quality of Life model (SQOL) as "directional intra-human" (i.e., QOL components defined as the conservative and expressive modes of the social and of the personality subsystems). Indeed, the study's substantive conclusions hinge on this finding.

The faceted external variable procedure has an important desideratum: the conclusions of the analysis are *invariant* under reversal of the direction of the external-variable range. Since the external variable bears no allegiance to the content universe, there is no criterion for deciding which side of its range, the top or the bottom, should be considered "up" for computing correlations with content-universe variables. Indeed, if its range is reversed, the computed correlations will all have the opposite sign, and the meaning of the external variable would be reversed, too. But the region(s) identified by the faceted external variable procedure would be the same and so would be the substantive conclusions drawn.

Cross-References

Systemic Quality of Life model (SQOL)

Faceted Action System Theory (FAST)

Facet Theory

SSA

References

Benish-Weisman, M. & Shye, S. (2011). Life quality of Russian immigrants to Israel: patterns of success and of unsuccess. *Social Indicators Research* Volume 101, Number 3, 461-479.

Borg, I., Groenen, P. (2005). *Modern Multidimensional Scaling: theory and applications* (2nd ed.). New York: Springer-Verlag. pp. 207–212. ISBN 0387948457.

Borg, I. and Shye, S. (1995). *Facet Theory: Form and Content*. Thousand Oaks CA: Sage.

Denesh, I. and Shye, S. (1993). Facet-inducing external variables: a new procedure for the FSSA. Proceedings of the Fourth International Facet Theory Conference, Prague, 1993.

Foa, U. G. (1965). New developments in facet design and analysis. Psychological Review, 72 (4) 262-274.

Guttman, L. (1968). A General nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrica*, 33, 469-506.

Guttman, L. (1982). Facet Theory, Smallest Space Analysis, and Factor Analysis. *Perceptual and Motor Skills, 54*, 491-493. (Addendum to Guttman, R. and Shoham, I. (1982). The structure of spatial ability items: a faceted analysis. *Perceptual and Motor Skills, 54*, 487-493).

Kingsley R. (1991). *Description of the Routines for Separation of Points in the Square into Separate Regions.* Jerusalem: The Louis Guttman Israel Institute of Applied Social Research. (Unpublished Technical Report.)

Kruskal, J. B. (1964). Multidimensional Scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrike, 29*, 1-27.

Lingoes, J. C. (1973). *The Guttman-Lingoes Nonmetric Program Series*. Ann Arbor, MI: Mathesis Press.

Shye , S. (1971). *A note on mapping sentences and their derivation through the use of SSA (Technical Report).* Jerusalem: The Louis Guttman Israel Institute of Applied Social Research. (Also in Abu Gosh, S. & Shye, S. (1971). *The impact of the elective local council system on Arab traditional society in Israel Vol.III*. Jerusalem: The Louis Guttman Israel Institute of Applied Social Research).

Shye, S. (Ed.). (1978a). *Theory Construction and Data Analysis in the Behavioral Sciences*. San Francisco: Jossey-Bass.

Shye, S. (1978b). Achievement motive: a faceted definition and structural analysis. *Multivariate Behavioral Research, Vol. 13*, No.3.

Shye, S. (1985). *Multiple Scaling: The Theory and Application of Partial Order Scalogram Analysis.* Amsterdam: North-Holland.

Shye, S. (1988). Inductive and Deductive Reasoning: A Structural Reanalysis of Ability Tests. *Journal of Applied Psychology, 73*, pp. 308-311. (Appendix: Multidimensional Scaling Versus Factor Analysis: A Methodological Note).

Shye, S. (1989). The Systemic Life Quality Model: A Basis for Urban Renewal Evaluation. *Social Indicators Research, 21*, pp. 343-378. Stable URL: http://www.jstor.org/stable/27520775

Shye, S. (1991). *The integration of Research Contents and Topological Spaces: Confirmatory SSA. (A new program for the PC).* The Third International Facet Theory Conference. Jerusalem, 1991.

Shye, S. (1998). Modern Facet Theory: Content design and measurement in behavioral research. *European Journal of Psychological Assessment, 14,* (2), 160-171.

Shye, S. (1999). Facet Theory. *Encyclopedia of Statistical Sciences, Update, Vol. 3*. New York: Wiley, 231-239.

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