UNIT 24 SCALOGRAM ANALYSIS

Structure

24.0 OBJECTIVES

After reading the unit, you should be able to:

- Explain the notion of measurement and its levels;
- Describe the concept of dimensionality;
- Explain the procedures for constructing indices and scales;
- Analyse One-dimensional scales and summated rating scales; and
- Discuss the method of scalogram analysis and its uses, particularly as a development planning technique, and also its limitations.

24.1 INTRODUCTION

This unit on scalorgram analysis follows naturally from the unit on preparation of data matrix (Unit 19). These two units are closest in relation to each other. One of the purposes of data theory is to provide a rationale for the use of scaling methods. This unit draws on the discussion of data theory, and describes the method of constructing scales, as also their uses. Other than a general discussion of scaling methods, the unit concentrates on one particular scale, the Guttman scale that is also known as scalogram. The unit also talks of the use of scalogram analysis in development planning. You have already read in Unit 23 about central place theory, which is a topic in regional science, the discussion on Central Place theory will help you understand the present unit in a better way. Often, in regional planning, scalogram is used as a technique. You will be familiarised with this important scaling method as also with some other important scaling techniques in this unit.

24.2 MEASUREMENT

The concept of measurement is basic to the activity of quantitative analysis. Whenever we attempt to qualify events or objects or even concepts we are engaged in measurement. But what do we exactly mean by measurement? Put simply, measurement means the assigning of numbers to events or objects. Let us try to make it a little more concrete. Suppose you consider five batsmen in cricket and take two characteristics, ability to play fast bowling and ability to play spin. You are now asked to estimate these two characteristics and assign 1 to the lowest level of a characteristic and 5 to the highest level of that characteristic. Now whatever the assessment in your mind, whatever the reason you choose for assigning a particular number to a particular level of a specific characteristic, you are engaged in measurement.

When we define measurement, we do not say anything about the quality of the measurement method, we are basically concerned with objects and events and with numerals to be assigned to these objects and events. A numeral is a symbol. It can be used to label objects. When we assign quantitative measuring to a numeral, it becomes a number.

We said earlier that we measure objects, but actually what we measure are the properties or characteristics of these objects. Even this is not strictly true. Actually what we measure are indicators or indicants of the properties of these objects. When the objects are physical, the indicants are simple. However, in social sciences or behavioural sciences, the indicants become much more complex and elusive.

From the indicants we arrive at the properties of the objects via operations that allow us to measure variables and constructs. Thus, we assign numerals to the behavioural indicants of properties. For example, when behavioural scientists are to measure intelligence, it is operationally defined on scores of intelligence tests.

The first and foremost step in any measurement exercise is an attempt at specifying the objects of the universe of discourse. This universe of discourse is also called the universal set. Once the universe of discourse is specified, we must define the property or properties of the objects in the universe of discourse. The collection of all the objects in the universe of this course is called the universal set. The next step in measurement is to classify the objects of the universal set according to some rule. These objects after classification are placed into subjects of the universal set.

Subsets are sets that are part of bigger sets. If a set B is a subset of a set A then every member of B is also a member of A, but not necessarily vice versa. For example, the set of horses is a subject of the set of four-footed animals. The set of North-Eastern States of India is a subject of the set of the states of India.

Putting the objects into subsets in such a way that every object of the universe of discourse belongs to some subject is called partitioning the universe of discourse. The subjects are mutually exclusive and non-exhaustive.

There are various levels and degrees of measurement. The strictest type of measure is a ratio-scale, where we can talk of an absolute zero and one level is say, twice or thrice of another. Obvious examples are weight and height. We can say one object is twice as long as another is. It does not matter whether we measure length in inches or centimetres. Similarly, we can say that one object is twice as heavy as another is, whether we measure in kilograms or pounds.

Somewhat less strict a measure is the interval scale. Interval scales do not have a notion of absolute zero. Ratios make no sense. Temperature is an example. We cannot say that water at 100^0 C is twice as hot as water at 50^0 C, because the corresponding temperature measured in Fahrenheit would not be trouble of each other. However, ratios of differences do make sense. The gap between 100^0 C and 90^0 C and 85^0 C, even when we measure in Fahrenheit scale is the same (the corresponding temperatures are 212^0 F, 194^0 F, and 185^0 F. (In the Celsius scale, the differences are 10 and 5 whereas in the Fahrenheit it is 18 and 9. This means the differences are not the same in the two scales but one difference is twice that of the other in both scales – 10 is twice that of 5 and 18 twice that of 9). Interval and ratio level measures allow us to make exact determinations of the distances between two points; they have fixed and equal intervals. For practical purposes of policy researchers, there is little difference between interval and ratio measures. Ratio scales have absolute and fixed zero points.

A scale may be of an ordinal measure type. These types are generally those where, for example, respondents are asked to rank their attitudes towards a specific issue. However, ordinal scales do not allow us to assume that there are equal intervals between each point in a given scale.

The nominal level is simply a classification. There is no sense of "order" in a nominal level of measurement. Nominal classifications are very precise in their ability to sort units into one category or another.

While nominal measures only classify, ordinal measurements imply order. In an ordinal measure, one can say that X category is more than or less than Y category. We can classify and order. A common ordinal measure is a scale. Ordinal scales do not allow us to assure that there are equal intervals between each point on our scale. A scale is a common example of an ordinal measure. Suppose we have a scale about a particular opinion in terms of somewhat agree, strongly agree…. and so on. It is ordinal measure.

Validity

A good measure should measure what it is supposed to measure. If we are measuring quality of service, we should not end up measuring quantity. This property of measuring what is supposed to be measured is called validity of the measurement technique. There are four validity tests:

(a) Face validity: To see valid at face value;

(b) Content validity: It means that the measure encompasses the totality of elements thought to be part of the concept that we are attempting to measure. No part of the concept should be left out of the measurement;

(c) Another test of validity is predictive validity; and

(d) A fourth test for validity is construct validity. Each item should measure the relevant particular concept relevant to the measurement.

The most important components of validity are sensitivity and specificity.

Reliability

A reliable measure is one, which, if applied time after time, will yield the same results (assuming no change in the thing to be measured). What is the relation between reliability and validity? A valid scale is always reliable, but a reliable scale is not always valid.

Measures should be chosen so as to be comprehensible to the potential audience. Furthermore, measures should also be chosen so as to minimise cost and be complete.

24.3 CONSTRUCTING INDICES AND SCALES

A scale is often an index. A researcher would typically obtain several different measures and combine them in an index. Those scales that are computed simply by adding together a number of items are sometimes called an index. Scales may be constructed from items asked in public opinion surveys, or from data found in public records, or generated from personal observations. In a simple scale,

responses to several items are merely added. But sometimes, in more complex scales, different items are given differing weights, depending on the relative importance of the items. Deciding on weight age to be given to each item is a complex process.

Given a particular concept that we are trying to measure, we must choose appropriate items in order to construct a relevant scale. A very simple type of scale is the Likert scale, which is an additive scale generally used in attitude measurement.

A good measure is complete; it adequately covers all-important aspects of the concept being measured. Often, this cannot be done with a single measure. In many circumstances, a researcher still wants to obtain several measures and combine them in an index, or scale. The simplest scale involves merely adding together several responses.

Choosing items

To choose items to construct a scale, we would be concerned with face validity and content validity. Some pertinent questions, which appear on surface include: Do all potential items we are considering seem to relate to the concept that we are trying to measure? Are all relevant aspects of the concept included in our scales?

Item Interrelationships

Next, we need to look at the interrelationship among the items. If all items are purporting to measure the same concept, they should be related to one another and to the overall scale. We can use cross tabular and correlation techniques. We also need to compare each individual item to the total scale score. This is called item analysis. If individual score is high on one item, but low on the whole scale, then that item is not measuring the same thing as the other items, and should be dropped.

Scoring the Scale

Followed by it is the step to score each respondent on the scale. This step involves deciding on weights, if any to assign to various responses, and then surveying each respondent's responses. After that, the respondents might be clustered in a few categories. Or the scale might need to be recoded in another form. Usually, extreme scale responses are a few in numbers.

Another problem in scoring the scale is to decide what to do with missing data. There are several ways to deal with this problem:

1. Throw the individuals with missing data out of the analysis;
2. Ignore the missing responses by assigning them zeroes;
3. Give each respondent who misses an item the average score in that item; and
4. Assign an individual with a missing item his own average score on the items he has answered.

Validating the Scale

A final step is building a scale is to validate it. There we might assess how well our scale relates to other variables, which are presumably related to it.
Likert Scales are additive scales generally employed in attitude measurement. Individuals are assessed on a 5.7-point scale of agreement. The procedures for constructing and scoring Likert scales are identical to those we have just discussed for scales in general.

Guttman Scaling

Scaling is the process of constructing a rule, which assigns symbols or numerals to individuals or their behaviour, for the purpose of measurement. One kind of scales is attitude scales, which help to measure the attitudes of respondents towards a particular thing.

There are three major types of attitude scales:

1. Summated rating scale;
2. Equal-appearing interval scales; and
3. Cumulative (or Guttman) scales.

Characteristic of summated rating scale

(i) All items are of equal value, or equally important. There is no scaling of items. The individuals responding to items are scaled. The scaling is done through the sum (or averages) of the individual s responses.

(ii) The intensity of the individual s responses.

Researchers often use interviews to obtain information from respondents. The interviewer asks the respondent a series of question to obtain answers to the research questions at hand. The interview may be structured (also called standardised) or it may be unstructured (unstandardised). Two types of schedule items (individual questions) are used in preparing interview schedules: fixed alternative (or closed schedules) and open ended.

A third type of items, which actually is a kind of fixed alternative item, is also used. These are scale items. A scale in this case counts of a group of verbal questions (items), which are administered to the respondents. An individual respondent expresses degree of agreement or disagreement/or some other response. Scale items have fixed alternatives. The respondent is placed at some point on the scale. This is used, for instance, in obtaining attitudes towards a specific thing.

Thus, a scale strictly speaking is a set of symbols or numerals constructed in a manner so that the individual can assign whatever is being measured by same rule. It is in this name, which has an interval measurement that ranks the particular concept at hand on some interval. Thus, scale becomes synonymous with measuring instrument. But for our purpose, a scale arises from the assigning of the items by the respondents. This scale is the measurement (ranking) of questionnaire items. So a scale is a broader concept, but we are concerned with the scales, which are tests like achievement lists or intelligence tests, or with scales that measure thing like attitudes. We will deal mainly with attitude scales.

There are broadly three types of attitude scales. Summated rating scales, equal appearing interval scales, and cumulative scales. A Likert scale is an example of the first type, the Thurstone scale is an example of the second type, and the Guttman Scale (or scalogram) is another name for the third type. We will presently elaborate a little on the Likert and Thurstone scales and discuss Guttman's scalogram analysis in detail, but here let us briefly discuss what these concepts mean.

A summated rating scale is a set of attitude items, with each item equally important, which are administered on the participants. Participants respond to each item with degrees (intensity) of agreement or disagreement. The scores of the responses for each individual are summed (average may be computed) to get that individuals attitude score. On the other hand, equal appearing interval scales, apart from assigning individual scores, also scale the attitude items. Each item is assigned a scale valve, and this value, shows the strength of an agreement response to the item. The third type of scale, the cumulative scale, consists of small set of homogeneous, One-dimensional items. A One-dimensional scale measures are variables only. The world 'cumulative' in the scales indicate that there is a cumulative relation between items and the total scores of individuals of the three, the summated rating scale focuses on the respondents and their rankings in the scale, the equal appearing interval scale focuses on the positions of the items on the scale. Cumulative scales focuses on the scalability of the items and the place of the individual on the scale.

Subjects and items are two important concepts in construction of scales. Usually subjects are individuals on whom the tests are administered. They are the respondents. Items are for the particular purpose, the different questions which the respondents answer. Each test may have several questionnaire items. The purpose of any item is permissibly to divide the given set of subjects into two or more mutually exclusive and exhaustive categories. Thus, suppose a mental test question is administered. Some of the respondents will answer it correctly and some incorrectly. Thus, the mental test question serves as an item, which helps to divide the group of respondents into those who answered it correctly and those who did not.

An important related idea is that of an indicator. Sometimes the researcher does not have measurement related to a certain concept, he/she tries to obtain observation on an indicator of the randomly concept. Now, the relation between an indicator and the underlying concept is often not a deterministic one, but a stochastic relation. There may arise problems of reliability; repeated measurement of the same indicator may give different values. Hence in practice, the researchers try to obtain as large a number of indicators as possible.

Certain problems arise when we try to relate the combined or indexed value of several different indicators to the underlying concept. As we shall see later in this unit, Guttman's suggestion was to look at the relationship between the indicators (items in this case) themselves.

Sometimes, the concept the researcher tries to measure, such as people's views about a certain policy, say, may be too abstract theoretically to lend itself to direct measurement. So the researcher tries to obtain a valid and reliable indicator of this abstract concept. In a manner of speaking, this empirical indicator may be called a scale. In many cases in the behavioural and social sciences, a scale is composed of an item or a set of items which are not administered to respondents in order that can be captured empirically. When

there is more than one item, these are combined to form a scale, the method of combination varying with the purpose for which the scale is being constructed.

Scaling may be needed to scale the respondents to scale stimulus or to scale both respondents and stimulus. For example, Likert scale is used to scale respondents. Similarly, only dents and stimuli are scaled, it is called response approach. The Scalogram, which we shall discuss extensively a little later, is an example of response approach.
Another important difference among scaling models relates to the type of data appropriate to the model. Recall that in the unit on Data Matrix, we had discussed Coomb's theory of data. To recapitulate briefly, there are four kinds of data. Preferential choice data concerned with the ranking of stimuli according to some criterion or purpose. Single stimulus data involve the subject (respondent) responding to ranking or comparison of stimulus. Question with 'yes' or 'no' response would be indicative of this type of data. Likert scale involves this kind of data. A third type of data is stimulus comparison data. There are the respondents in a given a set of stimuli asked to choose those that more or less possess characteristics that others have. This involves a stimulus-oriented and not respondent oriented approach. The final type of data suggested by counts is similarities data. All pairs of stimuli are formed, and respondents are asked to put the most similar attributes into pairs.

A final characteristic in terms of which scaling models can be distinguished is whether the models are One-dimensional. Dimensionality is a complex concept and means different things in different scaling models. Recall the discussion on dimensionality presented earlier. For scaling purposes, we can state that One-dimensional scales employ techniques which aim that the researcher can demonstrate to empirically correspond to a single socio-psychological dimension. Supposing the researcher feels that there is a single dimension underlying a set of observation than the researcher would use a One-dimensional scale. Psychologists use it to study a particular personality trait. In contrast, multidimensional scaling models explicitly allow for the possibility that there is more than one dimension underlying a set of observations. Multidimensional scaling methods can be used to determine the existence of a single latent dimension underlying a set of observations. Since One-dimensional scaling methods are simpler than multidimensional methods, and since in many multidimensional scaling methods, a One-dimensional method is a necessary prior step, many experts recommend a greater use of One-dimensional scale. Also, One-dimensional scaling uses concepts, which have a clear correspondence with concepts in social sciences. It is easier to analyse and construct concepts using One-dimensional methods.

24.4 AN INTRODUCTION TO ONE-DIMENSIONAL SCALING

To get information about people's opinion, the policy makers may devise a multiple-item questionnaire. There are several ways of combining these items into a scale, depending on the use to which the resulting scale is to be put. Each of these methods is related to a specific scaling model. A scale is a technique to develop a new measure, a technique that is internally consistent. Scaling is the process of constructing a rule, which assigns symbols or numerals to individuals or their behaviour, for the purpose of measurement.

Scaling models largely have these distinct uses. First, scaling models may help in testing a hypothesis. In unit 19, we touched upon hypothesis testing in our

discussion of research methods. Secondly, scaling can be used merely to describe a data structure that is to uncover latent or hidden dimensions that underline a set of obtained observations. Finally, scaling can be used to develop a scale on which individuals can be scored.

The basic idea of scaling is to represent objects as points within a space. All scaling methods are common in the sense that the dimensions of the space correspond to the sources of variation underlying the objects. But the way in which one proceeds from observation to data, and then to geometric representation differs from scale to scale.

Three methods of metric scaling are as follows:

1. Principal component analysis;
2. Multidimensional preference scaling; and
3. Correspondence analysis.

The basis of scalogram analysis is the scalogram. Scalogram analysis is a technique to examine whether a set of items is consistent, in the sense that they all measure the same thing. If all items measure the same thing, they are said to be One-dimensional, that is, they have a single dimension. Recall the meaning of the concept of dimensionality given in the unit on data matrix.

24.5 GUTTMAN SCALE (SCALOGRAM)

In 1944, sociologist Lous Guttman introduced the Scalogram. Scalogram is also called cumulative scaling. He put forward the salogram as a procedure designed to order both items as well as the subjects (the persons whose characteristics we want to scale) with respect to some underlying cumulative dimension. It is a deterministic model of scaling. This means that each value of the scale is a single number desired from the underlying continuum. Guttman continued to work on this and in work done in the 1950s, suggested ways of dealing with those scales for which no criterion for validating them is available. Basically, Guttman suggested that when such criteria are absent, we must look at the relations between the items themselves.

Historically, Guttman scaling was developed as a critical alternative to Thurstone and Likert methods of scaling, particularly attitude scaling. Guttman particularly stressed the point that neither Likert's nor Thurstone's techniques allowed one to determine whether a series of items belong on a One-dimensional continuum. Traditionally, researchers used item analysis methods to build a scale. In other words, items were chosen for their ability to predict a total score. Guttman suggested that to establish that an item is part of a single underlying dimension, we have to see if a scale is able to predict responses to all of its component items on the basis of total scores. Thus, the basic difference between Guttman's and the other approaches to scaling lies in differences in the conceptualisation of dimensionality.

Scalogram analysis is a technique to examine whether a set of items is consistent, in the sense that they all measure the same thing. If all items measure the same thing, they are said to be One-dimensional, that is, they have a single dimension.

To do scalogram analysis, one looks at the items, and sees whether the responses to these items conform to the scalogram. But the scalogram does not tell us whether the items measure the construct that we wish to measure; it merely tells us whether the item measures the same thing. Thus, it is a method to test for consistency, not validity. Usually, the starting point is a set of items in which one is interested because we feel that they measure a particular behavioural or psychological construct (for example ability, attitude). Thus the items are considered as an operational definition of a psychological construct. However, once the scalogram is applied to these items, it does not matter which psychological construct they are related to. They are operational definitions. Instead the items themselves assume interest, and the crucial question is whether the responses to these items conform to the scalogram.

What is the notion of scale in the scalogram analysis? Guttman argued that given a population of objects (what we called the universe of discourse), there would be a group of attributes of these objects. This group will display a multivariate frequency distribution of these attributes. If from the multivariate frequency distribution a quantitative variable can be derived, which can channel the objects in such a way that each attribute is a simple function of that quantitative variable. This is called quantitative variable. The multivariate function is itself called a scale.

The ideal deterministic model is one where there is a perfect relationship between scale score and item score. Scalogram analysis helps in anticipating that there might be a deviation of the actual response patterns from those required by the ideal model.

Let us pause here awhile to go over the sense of scalogram analysis. Strictly speaking, scalogram analysis is not a method for constructing or developing an attitude scale. Rather, it is a process to determine whether a series of items and sample of subjects conform to a set of specified criteria which has been taken to be the requirement of a Guttman scale. Thus, Guttman scaling is here discussed from a hypothesis – testing standpoint. Sometimes, however, the Guttman scaling model is used as an exploratory method to select items that conform to scale criteria from a longer set of items.

Let us now look a little at the kind of use to which scalogram analysis has been put for decentralised planning. We have seen in an earlier unit that spatial aspects come to the fore in decentralised planning. For decentralised planning one of the key concepts is the idea of ‘central places’. Now according to Christaller’s central place theory, there is a hierarchy of central settlements. These settlements can be graded and ranked using scalogram technique. A proper scalogram analysis can provide insight into the requirement and specialisation of various settlements in terms of socio-economic functions. At each level in the hierarchy a proper central place can be determined using scalogram technique. Settlements can be ranked using Guttman’s scalogram analysis in terms of institutions, facilities or infrastructure present. Thus scalogram is an important technique in spatial planning, decentralised development policy and rural development. It is part of a group of useful tools for planners, such as input-output analysis, mix-and share analysis, etc

24.6 ACTIVITY

1. Distinguish between One-dimensional and multidimensional scaling processes.

2. Provide an example of Likert scaling process
3. What way would you use scalogram if you were planning a township or a cluster of urban settlements that were surrounded by rural spaces? Discuss with examples.

24.7 CONCLUSION

This unit dealt with issues in measurement and the construction of indices and scales. The unit began by explaining the notion of measurement and it's various types: nominal, ordinal, interval and ratio. It discussed the important concepts of validity and reliability of measurement. The unit then went on to discuss the idea of scales and indices, talking about various ways in which items can be combined to form scales.

The unit explained the notion of dimensionality and items, and named some important scales like Liket and Thurstone scales. It described cumulative scales and summative scales, and distinguished between One-dimensional and multidimensional scales. It discussed in detail the scalogram technique propounded and developed by Guttman, and mentioned some uses of this scaling process in decentralised planning.

24.8 REFERENCES AND FURTHER READINGS

"International Bank for Reconstruction and Development, Rural Development Sector Policy Paper", World Bank, Washington, 1975.

Rondinelli, Dennis A. and Kenneth Ruddle, Urbanisation and Rural Development: A Spatial Policy for Equitable Growth, Praeger Publishers, New York, 1978.

Johnson, E.A.J., The Organisation of Space in Developing Countries, Harvard University Press, Cambridge, 1970.