Construction of Scales

Content Outline

1. Objectives
2. Introduction
3. Thurstone Scale
   3.1. Steps for creating a Thurstone scale
4. Likert Scale
   4.1. Steps for Constructing a Likert Scale
5. Guttman Scale
   5.1. Steps for creating Guttman Scale
6. Semantic Differential Scale
   6.1. Steps for Creating Semantic Differential Scale
7. Do’s and Don’ts
8. Summary

1. OBJECTIVES
   At the end of this session you will be able to
   ✓ Create Thurstone scale
   ✓ Create Likert Scale
   ✓ Create Guttman Scale
   ✓ Create Semantic Differential Scale

2. Introduction
   There are different types of scales used for data collection. They are generally classified into 4 types, Thurstone Scale, Likert Scale, Guttman Scale, and Semantic Differential Scale. In Unit on .... their characteristics, their use for the researcher is discussed. In this unit we will discuss the process of construction of these scales.
3. **Thurstone Scale**

In Psychology and Sociology, the **Thurstone Scale** was the first formal technique to measure an attitude. It was developed by Louis Leon Thurstone in 1928, as a means of measuring attitudes towards religion. It is made up of statements about a particular issue, and each statement has a numerical value indicating how favorable or unfavorable it is judged to be. People check each of the statements to which they agree, and a mean score is computed, indicating their attitude (Wikipedia).

Thurstone suggested 3 different methods for preparing the scales. He invented 3 different methods for developing uni-dimensional scales (Knowledge Base, 2006). They are:

- method of equal appearing intervals
- method of successive interval
- method of paired comparisons

There are scales associated with statistical analysis and these are nominal scale, ordinal scale, interval scale and ratio scale. While preparing the attitude scale we can assume that interval scale is appropriate for this scale. The respondents will be able to respond to the stimuli if the scale is on interval scale. He developed the procedure by which 20 statements (some positive and some negatives) can be generated. The respondents will respond to these statements, they would agree or they would not agree but they would put a tick mark. It is not 5 point scale it is only Agree or Disagree. While scoring it one will have to be careful: for scoring positive and negative statements appropriately.

### 3.1 Steps in construction of Thurstone Scale

For developing a Scale, the first thing is to refer to the objective of the study. The researcher also has to refer to the theoretical framework on the basis of which the objectives were prepared. The researcher has already identified the concept or the idea or a trait for investigation.

**Step 1**

Now the first step is to collect many statements which are related to these concepts. These statements can be 100-200 in number. The characteristic of these statements is that they are non-factual; they should not be giving the facts. For e.g.; - We shouldn’t take a statement like “Delhi is a capital of India”. It is a factual statement, that’s not an opinion. The opinion based statements should be selected. So if a statement is “Delhi, the Capitol of India, is a beautiful city” expresses an opinion and can therefore be included. So one has to select, collate and collect the statements which are opinion based and not factual.
**Step 2**

The second step is to get them rated. This can be done by requesting a pool of judges. This number could be from 50-300, a large number of people are invited to work as judges to rate these statements. 11 categories are formed ranging from negative or non-favorable attitude to positive or favorable attitude. Now the judges are told to rate each statement on these 11 categories, they will put a tick mark on one of them. So it is **Negative----- Neutral----- Positive** continuum. At the end of this second step, all the judges will have rated all the 120-150 statements which have been collected.

**Step 3**

Now that the judges have rated all those statements the third step is to give a Median Ranking. Every statement is now rated. Median Value is the center-point around which 50% cases on both sides lie. The statements that have too broad variations, should be removed. Because if people are rating a statement from 1 to 11 that means it is really not correctly understood. They must be ambiguous or people must be thinking that they are irrelevant, so such statements should be discarded. From among 120 to 200 statements, only 20 statements are selected. Ambiguous, irrelevant statements, those which have wide variations are removed. Once 20 statements are identified, opinionnaire or attitude scale is ready.

This attitude scale is administered to 200-300 respondents (representing a sample, drawn from the same population) and they are asked to rate “agree or disagree”. Each statement has a positive value or negative value. This is still a process of construction of Thurstone scale. This step is to test that these 20 statements are correctly selected.

**Step 4**

Once it is established that these 20 statements are correctly selected, the Attitude Scale is ready.

**4. Likert Scale**

Likert Scale is widely used in the field of research, especially in psychology, sociology, and education. Likert Scale has a list of statements and a 5 point scale. Three point, 7 point or 9 point scale can also be used.

**Step 1**

The first step in constructing Likert Scale is to collect as many statements related to the concept as possible. These will be 100-200 statements either they are favorable
or they are not favorable to the concept. These statements clearly indicate that they are favored or not favored.

**Step 2**
The second step is to administer these statements to a sample which is drawn from the same population. This sample is also representative of the same sample which will be used for the research study. So you invite this sample, you select this sample and give the treatment to them to rate. Now you have to score this most favorable statements will get the highest score because you have selected the statements which are either most favorable or least favorable. On receiving the rating for all the statements, the statements which are rated with a wide variation are discarded. The statements which differentiates the best between the negative and positive are selected. They can be between 6 and 20. These statements are differentiating i.e. they can be very clearly rated by the sample either favorable or not favorable. In this list there are items which are positive and there are items which are negative.

**Example**
Measuring Student Attitudes toward Learning with Social Media: Validation of the Social Media Learning Scale (Knezek, Mills and Wakefield, 2012)
The researchers developed this scale called Social Media Learning (SML) Scale. They used 147 adult social media users to validate the Likert type scale. There were only 8 items and there were 5 categories: Strongly Agree to Strongly Disagree through Neutral. Cronbach’s Alpha was calculated which was 0.78 and is quite high.

**5. Guttmann Scale**
Let us see the third type of scale Guttmann scale, how to construct Guttmann scale.

**Step 1**
The researcher selects statements that are applicable to the measurable objectives. The objectives are already identified, the concept or a trait, or a thought, the attitude towards which is investigated.

**Step 2**
The researcher selects about 100 judges to test the statements and a small list of statements is culled out of the big list.
**Step 3**
The statements which are 80% and above in agreement or disagreement are discarded, because they are the extreme.

**Step 4**
The selected statements are arranged from most favorable to least favorable. These statements are arranged from left to right, from most favorable to least favorable.

**Step 5**
The statement which failed to discriminate between these two extremes are discarded.

**Step 6 : Calculation of Coefficient of Reproducibility**
1. Then calculate the coefficient of reproducibility.
2. First calculate the number of errors, favorable responses that do not fit the pattern.
3. Coefficient or reproducibility = 1- (number of errors/total number of responses).
4. If reproducibility is equal to 0.9 or greater a uni-dimensional scale is said to exist, this Scale can be accepted.

6. **Semantic Differential Scale**

The Semantic Differential refers to the words which are on the scale to two extremes, one is positive another is negative. For example, Lazy and Industrious. This refers to the respondents attaching meaning to that word. This meaning is attached by the respondents to that word. Meaning is of two types and that's why the words can be of two types, one is **denotative** and other is **connotative**. Denotative means denoted. If we say “Horse” it is a name of an animal, so that is the meaning given to that word. This is denotative. Connotative is not the meaning given but the meaning is attached to that word. The meaning attached to the word "Horse" is power. This becomes a connotative meaning.

The Semantic Differential uses this kind of connotative meaning. It asks the respondent the meaning they wish to attach to the pair of words. Connotative meaning is not the actual explicit meaning which that word has. It is the meaning attached to that word by the respondent. There are polar words or polar adjectives, e.g. good and bad, beautiful and ugly these are the two extreme words but they are polar adjectives. Osgood first suggested this type of scale which is called semantic differential scale. He used three areas: the strength, value and activity. He thought
that any scale prepared can have these three not necessarily all three together. Each one of them have a peculiar place in the scale.

In case of a ‘Strength’ a pair of polar adjectives ‘strong and weak’ can be used. Similarly ‘decisive and indecisive’ pair can be used. ‘Value’ can be ‘good and bad’ and ‘cheap and expensive’. For the ‘Activity’ one can use the word ‘active and passive’ that shows how actively people are engaged into that event or process. ‘Lazy and industrious’ is the pair which gives the sense of activity. Osgood suggested some of these adjectives which could be used as polar pair.

**Construction of Semantic Differential Scale**

Now let us see how semantic differential scale is prepared.

**Step 1**

The first step is to refer to theoretical framework, refer to objectives, and identify the concepts which one wants to investigate. Now select pairs which would define, explain these concepts. Identify or select the appropriate word which describes the concepts. You may be having two groups: experimental and control. Give these words to them and ask them which words clearly define, clearly explain the selected concepts.

Ask them to rate the list of 25-30 adjectives, for their appropriateness, words which are appropriate and which are not so appropriate. The ranking will show that the highest ranking words are relevant, more appropriate. The respondents must be given clear-cut instructions what is expected of them.

They need to remember that these words are not denotative but are connotative so the meaning attached to the word is important rather than the actual/explicit meaning of the word.

Once you get those ratings from two different groups with different instructions, find out the pairs clearly coming out, are distinguishing, differentiating from others. After analysis of the data from these raters, short-list words for the scale, not more than 12-14 statements. There need not be 50 statements in scale. There can be 8-12. Semantic Differential Scale is a short scale.

Once the scale is ready, do the pilot testing, by selecting sample which resembles or represents the population but not the sample which will be used in real study. Analyze the data and establish its validity and reliability.
DO’s and DON’TS for preparing Semantic Differential Scale

Do’s
What is more important is clear cut instructions to the respondent about the expectations of researcher to respond. Generally in semantic differential scale a 7-point scale is used but it can be a 5 point or 3 point scale as well.

Don’ts
There will not be more than 20 statements, 12 can be optimal, better to use fewer than 12.

Don’t put all good on one side and all bad on other side. That is positive on one side and negative on other side. That will create monotony and respondents may go on rating one particular point. In order to avoid that mix positive and negative in both the lines.

7. Summary
We have seen how to construct various scales as they are important tools for data collection. If they are faulty, naturally that will show in the data collection and the results will be affected. So it is necessary to understand the correct procedure of constructing these scales and using them.