Chapter 3: Levels Of Measurement And Scaling

A common feature of marketing research is the attempt to have respondents communicate their feelings, attitudes, opinions, and evaluations in some measurable form. To this end, marketing researchers have developed a range of scales. Each of these has unique properties. What is important for the marketing analyst to realise is that they have wildely differing measurement properties. Some scales are at very best, limited in their mathematical properties to the extent that they can only establish an association between variables. Other scales have more extensive mathematical properties and some, hold out the possibility of establishing cause and effect relationships between variables.

Chapter Objectives

This chapter will give the reader:

- An understanding of the four levels of measurement that can be taken by researchers
- The ability to distinguish between comparative and non-comparative measurement scales, and
- A basic tool-kit of scales that can be used for the purposes of marketing research.

Structure Of The Chapter

All measurements must take one of four forms and these are described in the opening section of the chapter. After the properties of the four categories of scale have been explained, various forms of comparative and non-comparative scales are illustrated. Some of these scales are numeric, others are semantic and yet others take a graphical form. The marketing researcher who is familiar with the complete tool kit of scaling measurements is better equipped to understand markets.

Levels of measurement

Most texts on marketing research explain the four levels of measurement: nominal, ordinal, interval and ratio and so the treatment given to them here will be brief. However, it is an important topic since the type of scale used in taking measurements directly impinges on the statistical techniques which can legitimately be used in the analysis.

Nominal scales

This, the crudest of measurement scales, classifies individuals, companies, products, brands or other entities into categories where no order is implied. Indeed it is often referred to as a categorical scale. It is a system of classification and does not place the entity along a continuum. It involves a simply count of the frequency of the cases assigned to the various categories, and if desired numbers can be nominally assigned to label each category as in the example below:

Figure 3.1 An example of a nominal scale

Which of the following food items do you tend to buy at least once per month? (Please tick)		
Okra	Palm Oil	Milled Rice
Peppers	Prawns	Pasteurised milk

The numbers have no arithmetic properties and act only as labels. The only measure of average which can be used is the mode because this is simply a set of frequency counts. Hypothesis tests can be carried out on data collected in the nominal form. The most likely would be the Chi-square test. However, it should be noted that the Chi-square is a test to determine whether two or more variables are associated and the strength of that relationship. It can tell nothing about the form of that relationship, where it exists, i.e. it is not capable of establishing cause and effect.

Ordinal scales

Ordinal scales involve the ranking of individuals, attitudes or items along the continuum of the characteristic being scaled. For example, if a researcher asked farmers to rank 5 brands of pesticide in order of preference he/she might obtain responses like those in table 3.2 below.

Figure 3.2 An example of an ordinal scale used to determine farmers' preferences among 5 brands of pesticide.

Order of preference	Brand
1	Rambo
2	R.I.P.
3	Killalot
4	D.O.A.
5	Bugdeath

From such a table the researcher knows the order of preference but nothing about how much more one brand is preferred to another, that is there is no information about the interval between any two brands. All of the information a nominal scale would have given is available from an ordinal scale. In addition, positional statistics such as the median, quartile and percentile can be determined.

It is possible to test for order correlation with ranked data. The two main methods are Spearman's Ranked Correlation Coefficient and Kendall's Coefficient of Concordance.

Using either procedure one can, for example, ascertain the degree to which two or more survey respondents agree in their ranking of a set of items. Consider again the ranking of pesticides example in figure 3.2. The researcher might wish to measure similarities and differences in the rankings of pesticide brands according to whether the respondents' farm enterprises were classified as "arable" or "mixed" (a combination of crops and livestock). The resultant coefficient takes a value in the range 0 to 1. A zero would mean that there was no agreement between the two groups, and 1 would indicate total agreement. It is more likely that an answer somewhere between these two extremes would be found.

The only other permissible hypothesis testing procedures are the runs test and sign test. The runs test (also known as the Wald-Wolfowitz). Test is used to determine whether a sequence of binomial data - meaning it can take only one of two possible values e.g. African/non-African, yes/no, male/female - is random or contains systematic 'runs' of one or other value. Sign tests are employed when the objective is to determine whether there is a significant difference between matched pairs of data. The sign test tells the analyst if the number of positive differences in ranking is approximately equal to the number of negative rankings, in which case the distribution of rankings is random, i.e. apparent differences are not significant. The test takes into account only the direction of differences and ignores their magnitude and hence it is compatible with ordinal data.

Interval scales

It is only with an interval scaled data that researchers can justify the use of the arithmetic mean as the measure of average. The interval or cardinal scale has equal units of measurement, thus making it possible to interpret not only the order of scale scores but also the distance between them. However, it must be recognised that the zero point on an interval scale is arbitrary and is not a true zero. This of course has implications for the type of data manipulation and analysis we can carry out on data collected in this form. It is possible to add or subtract a constant to all of the scale values without affecting the form of the scale but one cannot multiply or divide the values. It can be said that two respondents with scale positions 1 and 2 are as far apart as two respondents with scale positions 4 and 5, but not that a person with score 10 feels twice as strongly as one with score 5. Temperature is interval scaled, being measured either in Centigrade or Fahrenheit. We cannot speak of 50°F being twice as hot as 25°F since the corresponding temperatures on the centigrade scale, 10°C and -3.9°C, are not in the ratio 2:1.

Interval scales may be either numeric or semantic. Study the examples below in figure 3.3.

Figure 3.3 Examples of interval scales in numeric and semantic formats

Please indicate your views on Balkan Olives by scoring them on a scale of 5 down to 1 (i.e. 5 = Excellent; = Poor) on each of the criteria listed						
Balkan Olives are:						Circle the appropriate score on each line
Succulence	5	4	3	2	1	
Fresh tasting	5	4	3	2	1	
Free of skin blemish	5	4	3	2	1	
Good value	5	4	3	2	1	
Attractively packaged	5	4	3	2	1	
(a)						

Please indicate your views on Balkan Olives by ticking the appropriate responses below:					
	Excellent	Very Good	Good	Fair	Poor
Succulent					
Freshness					
Freedom from skin blemish					

Value for money							
Attractiveness of packaging							
(b)							

Most of the common statistical methods of analysis require only interval scales in order that they might be used. These are not recounted here because they are so common and can be found in virtually all basic texts on statistics.

Ratio scales

The highest level of measurement is a ratio scale. This has the properties of an interval scale together with a fixed origin or zero point. Examples of variables which are ratio scaled include weights, lengths and times. Ratio scales permit the researcher to compare both differences in scores and the relative magnitude of scores. For instance the difference between 5 and 10 minutes is the same as that between 10 and 15 minutes, and 10 minutes is twice as long as 5 minutes.

Given that sociological and management research seldom aspires beyond the interval level of measurement, it is not proposed that particular attention be given to this level of analysis. Suffice it to say that virtually all statistical operations can be performed on ratio scales.

Measurement scales

The various types of scales used in marketing research fall into two broad categories: comparative and non comparative. In comparative scaling, the respondent is asked to compare one brand or product against another. With noncomparative scaling respondents need only evaluate a single product or brand. Their evaluation is independent of the other product and/or brands which the marketing researcher is studying.

Noncomparative scaling is frequently referred to as monadic scaling and this is the more widely used type of scale in commercial marketing research studies.

Comparative scales

Paired comparison[2]: It is sometimes the case that marketing researchers wish to find out which are the most important factors in determining the demand for a product. Conversely they may wish to know which are the most important factors acting to prevent the widespread adoption of a product. Take, for example, the very poor farmer response to the first design of an animal-drawn mould board plough. A combination of exploratory research and shrewd observation suggested that the following factors played a role in the shaping of the attitudes of those farmers who feel negatively towards the design:

- Does not ridge
- Does not work for inter-cropping
- Far too expensive
- New technology too risky
- Too difficult to carry.

Suppose the organisation responsible wants to know which factors is foremost in the farmer's mind. It may well be the case that if those factors that are most important to the farmer than the others, being of a relatively minor nature, will cease to prevent widespread adoption. The alternatives are to abandon the product's re-development or to completely re-design it which is not only expensive and time-consuming, but may well be subject to a new set of objections.

The process of rank ordering the objections from most to least important is best

approached through the questioning technique known as 'paired comparison'. Each of the objections is paired by the researcher so that with 5 factors, as in this example, there are 10 pairs-

$$i.e \frac{n(n-1)}{2}, \frac{5(5-1)}{2} = 10 pairs$$

In 'paired comparisons' every factor has to be paired with every other factor in turn. However, only one pair is ever put to the farmer at any one time.

The question might be put as follows:

Which of the following was the more important in making you decide not to buy the plough?

| • The plough was too expensive | |
| • It proved too difficult to transport | |

In most cases the question, and the alternatives, would be put to the farmer verbally. He/she then indicates which of the two was the more important and the researcher ticks the box on his questionnaire. The question is repeated with a second set of factors and the appropriate box ticked again. This process continues until all possible combinations are exhausted, in this case 10 pairs. It is good practice to mix the pairs of factors so that there is no systematic bias. The researcher should try to ensure that any particular factor is sometimes the first of the pair to be mentioned and sometimes the second. The researcher would never, for example, take the first factor (on this occasion 'Does not ridge') and systematically compare it to each of the others in succession. That is likely to cause systematic bias.

Below labels have been given to the factors so that the worked example will be easier to understand. The letters A - E have been allocated as follows:

A = Does not ridge
B = Far too expensive
C = New technology too risky
D = Does not work for inter-cropping
E = Too difficult to carry.

The data is then arranged into a matrix. Assume that 200 farmers have been interviewed and their responses are arranged in the grid below. Further assume that the matrix is so arranged that we read from top to side. This means, for example, that 164 out of 200 farmers said the fact that the plough was too expensive was a greater deterrent than the fact that it was not capable of ridging. Similarly, 174 farmers said that the plough's inability to inter-crop was more important than the inability to ridge when deciding not to buy the plough.

Figure 3.4 A preference matrix

	A	B	C	D	E
A	100	164	120	174	180
B	36	100	160	176	166
C	80	40	100	168	124
D	26	24	32	100	102
E	20	34	76	98	100

If the grid is carefully read, it can be seen that the rank order of the factors is -

```
 Most important  E Too difficult to carry
                 D Does not inter crop
                 C New technology/high risk
                 B Too expensive
Least important  A Does not ridge.
```

It can be seen that it is more important for designers to concentrate on improving transportability and, if possible, to give it an inter-cropping capability rather than focusing on its ridging capabilities (remember that the example is entirely hypothetical).

One major advantage to this type of questioning is that whilst it is possible to obtain a measure of the order of importance of five or more factors from the respondent, he is never asked to think about more than two factors at any one time. This is especially useful when dealing with illiterate farmers. Having said that, the researcher has to be careful not to present too many pairs of factors to the farmer during the interview. If he does, he will find that the farmer will quickly get tired and/or bored. It is as well to remember the formula of n (n - 1)/2. For ten factors, brands or product attributes this would give 45 pairs. Clearly the farmer should not be asked to subject himself to having the same question put to him 45 times. For practical purposes, six factors is possibly the limit, giving 15 pairs.

It should be clear from the procedures described in these notes that the paired comparison scale gives ordinal data.

Dollar Metric Comparisons[3]: This type of scale is an extension of the paired comparison method in that it requires respondents to indicate both their preference and how much they are willing to pay for their preference. This scaling technique gives the marketing researcher an interval - scaled measurement. An example is given in figure 3.5.

Figure 3.5 An example of a dollar metric scale

Which of the following types of fish do you prefer?		How much more, in cents, would you be prepared to pay for your preferred fish?		
Fresh		Fresh (gutted)		$0.70
Fresh (gutted)		Smoked		0.50
Frozen		Smoked		0.60
Frozen		Fresh		0.70
Smoked		Fresh		0.20
Frozen(gutted)		Frozen		
From the data above the preferences shown below can be computed as follows:				
Fresh fish:	0.70	+ 0.70	+ 0.20	=1.60
Smoked fish:	0.60	+ (-0.20)	+ (-0.50)	=(-1.10)
Fresh fish(gutted):	(-0.70)	+ 0.30	+ 0.50	=0.10
Frozen fish:	(-0.60)	+ (-0.70)	+ (-0.30)	=(-1.60)

The Unity-sum-gain technique: A common problem with launching new products is one of reaching a decision as to what options, and how many options one offers. Whilst a company may be anxious to meet the needs of as many market segments as possible, it has to ensure that the segment is large enough to enable him to make a profit. It is always easier to add products to the product line but much more difficult to decide which models should be deleted. One technique for evaluating the options which are likely to prove successful is the unity-sum-gain approach.

The procedure is to begin with a list of features which might possibly be offered as 'options'

on the product, and alongside each you list its retail cost. A third column is constructed and this forms an index of the relative prices of each of the items. The table below will help clarify the procedure. For the purposes of this example the basic reaper is priced at $20,000 and some possible 'extras' are listed along with their prices.

The total value of these hypothetical 'extras' is $7,460 but the researcher tells the farmer he has an equally hypothetical $3,950 or similar sum. The important thing is that he should have considerably less hypothetical money to spend than the total value of the alternative product features. In this way the farmer is encouraged to reveal his preferences by allowing researchers to observe how he trades one additional benefit off against another. For example, would he prefer a side rake attachment on a 3 metre head rather than have a transporter trolley on either a standard or 2.5m wide head? The farmer has to be told that any unspent money cannot be retained by him so he should seek the best value-for-money he can get.

In cases where the researcher believes that mentioning specific prices might introduce some form of bias into the results, then the index can be used instead. This is constructed by taking the price of each item over the total of $ 7,460 and multiplying by 100. Survey respondents might then be given a maximum of 60 points and then, as before, are asked how they would spend these 60 points. In this crude example the index numbers are not too easy to work with for most respondents, so one would round them as has been done in the adjusted column. It is the relative and not the absolute value of the items which is important so the precision of the rounding need not overly concern us.
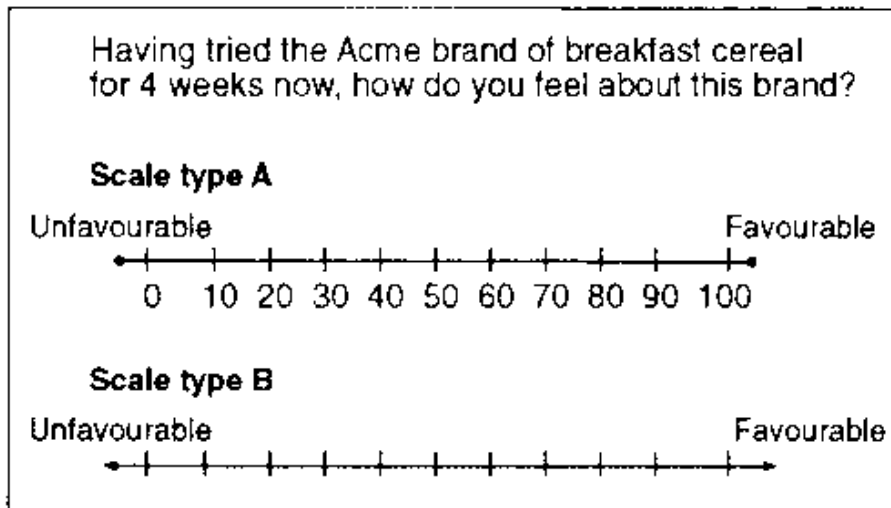
Figure 3.6 The unity-sum-gain technique

Item	Additional Cost ($s)	Index	Adjusted Index
2.5 wide rather than standard 2m	2,000	27	30
Self lubricating chain rather than belt	200	47	50
Side rake attachment	350	5	10
Polymer heads rather than steel	250	3	5
Double rather than single edged cutters	210	2.5	5
Transporter trolley for reaper attachment	650	9	10
Automatic levelling of table	300	4	5

The unity-sum-gain technique is useful for determining which product features are more important to farmers. The design of the final market version of the product can then reflect the farmers' needs and preferences. Practitioners treat data gathered by this method as ordinal.

Noncomparative scales

Continuous rating scales: The respondents are asked to give a rating by placing a mark at the appropriate position on a continuous line. The scale can be written on card and shown to the respondent during the interview. Two versions of a continuous rating scale are depicted in figure 3.7.

Figure 3.7 Continuous rating scales

Having tried the Acme brand of breakfast cereal for 4 weeks now, how do you feel about this brand?

When version B is used, the respondent's score is determined either by dividing the line into as many categories as desired and assigning the respondent a score based on the category into which his/her mark falls, or by measuring the distance, in millimetres or inches, from either end of the scale.

Whichever of these forms of the continuous scale is used, the results are normally analysed as interval scaled.

Line marking scale: The line marked scale is typically used to measure perceived similarity differences between products, brands or other objects. Technically, such a scale is a form of what is termed a semantic differential scale since each end of the scale is labelled with a word/phrase (or semantic) that is opposite in meaning to the other. Figure 3.8 provides an illustrative example of such a scale.
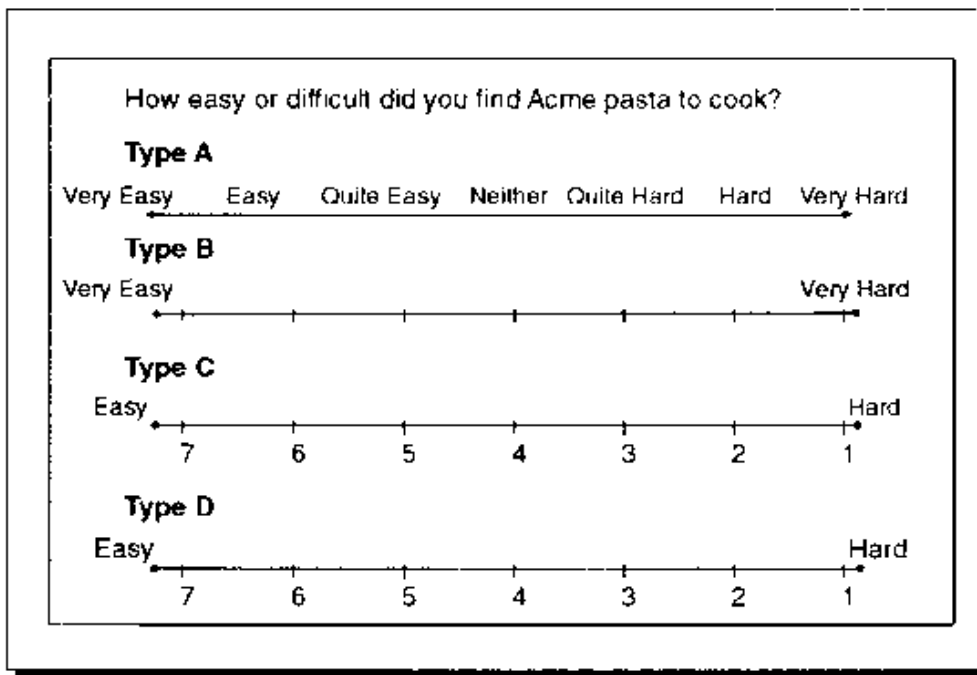
Consider the products below which can be used when frying food. In the case of each pair, indicate how similar or different they are in the flavour which they impart to the food.

Figure 3.8 An example of a line marking scale

For some types of respondent, the line scale is an easier format because they do not find discrete numbers (e.g. 5, 4, 3, 2, 1) best reflect their attitudes/feelings. The line marking scale is a continuous scale.

Itemised rating scales: With an itemised scale, respondents are provided with a scale having numbers and/or brief descriptions associated with each category and are asked to select one of the limited number of categories, ordered in terms of scale position, that best describes the product, brand, company or product attribute being studied. Examples of the itemised rating scale are illustrated in figure 3.9.

Figure 3.9 Itemised rating scales

How easy or difficult did you find Acme pasta to cook?

Type A

Very Easy   Easy   Quite Easy   Neither   Quite Hard   Hard   Very Hard

Type B

Very Easy   Very Hard

Type C

Easy   Hard
        7   6   5   4   3   2   1

Type D

Easy   Hard
        7   6   5   4   3   2   1

Itemised rating scales can take a variety of innovative forms as demonstrated by the two illustrated in figure 3.9, which are graphic.

Figure 3.10 Graphic itemised scales

Whichever form of itemised scale is applied, researchers usually treat the data as interval level.

Semantic scales: This type of scale makes extensive use of words rather than numbers. Respondents describe their feelings about the products or brands on scales with semantic labels. When bipolar adjectives are used at the end points of the scales, these are termed semantic differential scales. The semantic scale and the semantic differential scale are illustrated in figure 3.11.

Figure 3.11 Semantic and semantic differential scales
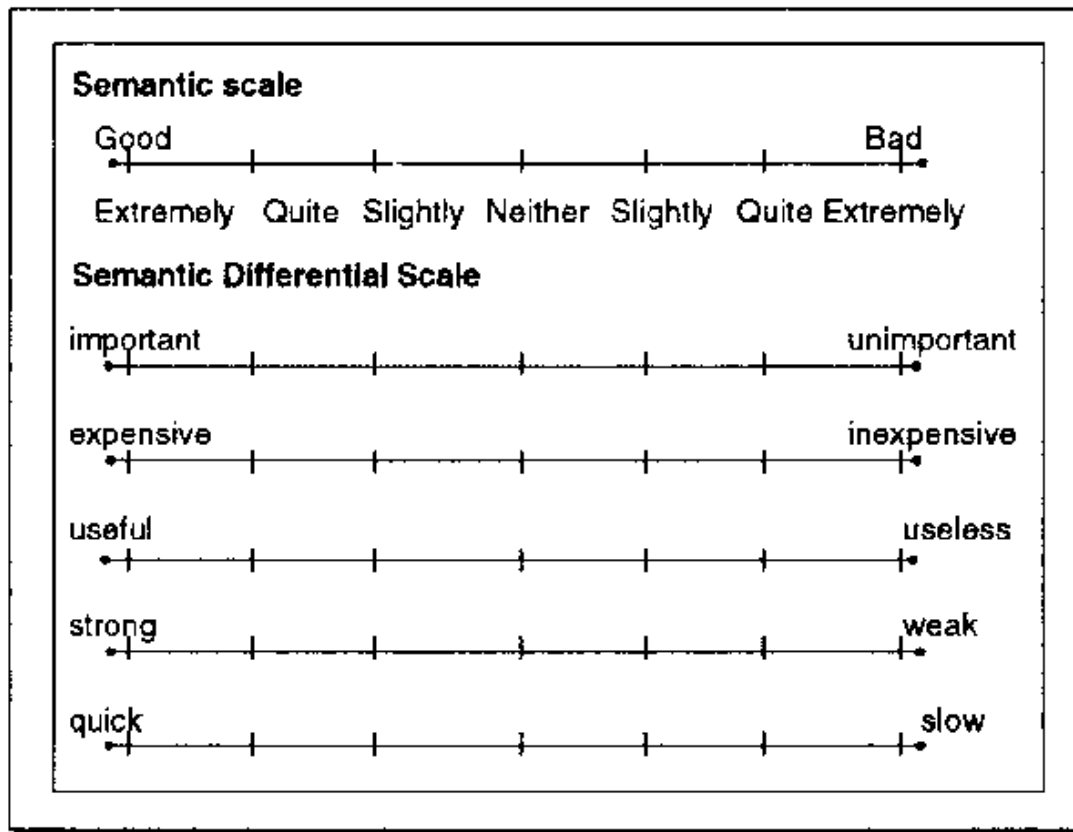
Likert scales: A Likert scale is what is termed a summated instrument scale. This means that the items making up a Liken scale are summed to produce a total score. In fact, a Likert scale is a composite of itemised scales. Typically, each scale item will have 5 categories, with scale values ranging from -2 to +2 with 0 as neutral response. This explanation may be clearer from the example in figure 3.12.

Figure 3.12 The Likert scale

	Strongly Agree	Agree	Neither	Disagree	Strongly Disagree
If the price of raw materials fell firms would reduce the price of their food products.	1	2	3	4	5
Without government regulation the firms would exploit the consumer.	1	2	3	4	5
Most food companies are so concerned about making profits they do not care about quality.	1	2	3	4	5
The food industry spends a great deal of money making sure that its manufacturing is hygienic.	1	2	3	4	5
Food companies should charge the same price for their products throughout the country	1	2	3	4	5

Likert scales are treated as yielding Interval data by the majority of marketing researchers.

The scales which have been described in this chapter are among the most commonly used in marketing research. Whilst there are a great many more forms which scales can take, if students are familiar with those described in this chapter they will be well equipped to deal

with most types of survey problem.

Chapter Summary

There are four levels of measurement: nominal, ordinal, interval and ratio. These constitute a hierarchy where the lowest scale of measurement, nominal, has far fewer mathematical properties than those further up this hierarchy of scales. Nominal scales yield data on categories; ordinal scales give sequences; interval scales begin to reveal the magnitude between points on the scale and ratio scales explain both order and the absolute distance between any two points on the scale.

The measurement scales, commonly used in marketing research, can be divided into two groups; comparative and non-comparative scales. Comparative scales involve the respondent in signaling where there is a difference between two or more producers, services, brands or other stimuli. Examples of such scales include; paired comparison, dollar metric, unity-sum-gain and line marking scales. Non-comparative scales, described in the textbook, are; continuous rating scales, itemised rating scales, semantic differential scales and Likert scales.

Key Terms

Comparative scales
Interval measures
Itemised scales
Line scales
Monadic scales
Nominal measures
Ordinal measures
Paired comparison
Ratio measures
Semantic differential
Unity-sum-gain

Review Questions

1. With which type of scale would Kendall's Concordance be used?

2. What is the more common name given to ordinal scales?

3. Why would a marketing researcher employ a dollar metric scale?

4. A researcher wants to measure consumer preference between 9 brands of vegetable oil and has decided to use the paired comparison method. How many pairs of brands will the researcher present to the respondents?

5. Explain what is meant by a semantic differential scale.

6. Two graphic scales are described in the textbook. What types of scale are these?

7. Look back to figure 3.5. What form of fish is most and least preferred?

8. What are the major statistical limitations to nominal scaled data?

Chapter References

1. Coombs, C. H. (1953).."Theory and Methods of Social Measurement", in *Research Methods in the Behavioral Sciences,* eds. Feslinger, L. and Ratz, D., Holt, Rinehart and

Winston.

2. Thurstone, L. L., (1927), "A Law of Comparative Judgment", *Psychological Review* 34, pp. 273-86.

3. Dillon, W. R., Madden, T. S and Firtle, N. H. (1994), *Marketing Research in a Marketing Environment,* 3rd edition, Irwin, p. 298.