

Likert Scales and Data Analyses

by I. Elaine Allen and Christopher A. Seaman

Surveys are consistently used to measure quality. For example, surveys might be used to gauge customer perception of product quality or quality performance in service delivery.

Likert scales are a common ratings format for surveys. Respondents rank quality from high to low or best to worst using five or seven levels.

Statisticians have generally grouped data collected from these surveys into a hierarchy of four levels of measurement:

1. **Nominal data:** The weakest level of measurement representing categories without numerical representation.
2. **Ordinal data:** Data in which an ordering or ranking of responses is possible but no measure of distance is possible.
3. **Interval data:** Generally integer data in which ordering and distance measurement are possible.
4. **Ratio data:** Data in which meaningful ordering, distance, decimals and fractions between variables are possible.

Data analyses using nominal, interval and ratio data are generally straightforward and transparent. Analyses of ordinal data, particularly as it relates to Likert or other scales in surveys, are not. This is not a new issue. The adequacy of treating ordinal data as interval data continues to be controversial in survey analyses in a variety of applied fields.^{1,2}

An underlying reason for analyzing ordinal data as interval data might be

the contention that parametric statistical tests (based on the central limit theorem) are more powerful than nonparametric alternatives. Also, conclusions and interpretations of parametric tests might be considered

Can this method be used as interval measures?

easier to interpret and provide more information than nonparametric alternatives.

However, treating ordinal data as interval (or even ratio) data without examining the values of the dataset and the objectives of the analysis can both mislead and misrepresent the findings of a survey. To examine the appropriate analyses of scalar data and when its preferable to treat ordinal data as interval data, we will concentrate on Likert scales.

Basics of Likert Scales

Likert scales were developed in 1932 as the familiar five-point bipolar response that most people are familiar with today.³ These scales range from a group of categories—least to most—asking people to indicate how much they agree or disagree, approve or disapprove, or believe to be true or false. There’s really no wrong way to build a Likert scale. The most impor-

TABLE 2 Likert Scale Example

Compared to face-to-face learning, outcomes from online learning are currently:			
	2003	2004	2006
Superior	0.6%	1%	1.8%
Somewhat superior	11.5%	10%	15.1%
Same	50.6%	50.6%	45%
Somewhat inferior	28.4%	28.4%	30.3%
Inferior	10.1%	10.1%	7.8%

Source: I. Elaine Allen and J.R. Seaman, "Making the Grade: Online Education in the United States," www.sloan-c.org, 2006.

tant consideration is to include at least five response categories. Some examples of category groups appear in Table 1.

The ends of the scale often are increased to create a seven-point scale by adding "very" to the respective top and bottom of the five-point scales. The seven-point scale has been shown to reach the upper limits of the scale’s reliability.⁴ As a general rule, Likert and others recommend that it is best to use as wide a scale as possible. You can always collapse the responses into condensed categories, if appropriate, for analysis.

With that in mind, scales are sometimes truncated to an even number of categories (typically four) to eliminate the "neutral" option in a "forced choice" survey scale. Rensis Likert’s original paper clearly identifies there might be an underlying continuous variable whose value characterizes the respondents’ opinions or attitudes and this underlying variable is interval level, at best.⁵

Analysis, Generalization To Continuous Indexes

As a general rule, mean and standard deviation are invalid parameters for descriptive statistics whenever

TABLE 1 Likert Scale Response Categories

Scale	1	2	3	4	5
	Never	Seldom	Sometimes	Often	Always
	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
	Most Important	Important	Neutral	Unimportant	Not important at all

data are on ordinal scales, as are any parametric analyses based on the normal distribution. Nonparametric procedures—based on the rank, median or range—are appropriate for analyzing these data, as are distribution free methods such as tabulations, frequencies, contingency tables and chi-squared statistics.

Kruskall-Wallis models can provide the same type of results as an analysis of variance, but based on the ranks and not the means of the responses. Given these scales are representative of an underlying continuous measure, one recommendation is to analyze them as interval data as a pilot prior to gathering the continuous measure.

Table 2 includes an example of misleading conclusions, showing the results from the annual Alfred P. Sloan Foundation survey of the quality and extent of online learning in the United States. Respondents used a Likert scale to evaluate the quality of online learning compared to face-to-face learning.

While 60%-plus of the respondents perceived online learning as equal to or better than face-to-face, there is a persistent minority that perceived online learning as at least somewhat inferior. If these data were analyzed using means, with a scale from 1 to 5 from inferior to superior, this separation would be lost, giving means of 2.7, 2.6 and 2.7 for these three years, respectively. This would indicate a slightly lower than average agreement rather than the actual distribution of the responses.

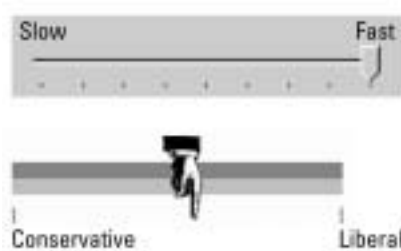
A more extreme example would be to place all the respondents at the extremes of the scale, yielding a mean of “same” but a completely different interpretation from the actual responses.

Under what circumstances might Likert scales be used with interval procedures? Suppose the rank data included a survey of income measuring \$0, \$25,000, \$50,000, \$75,000 or \$100,000 exactly, and these were measured as “low,” “medium” and “high.”

The “intervalness” here is an attribute of the data, not of the labels. Also, the scale item should be at least five and preferably seven categories.

Another example of analyzing Likert scales as interval values is when the sets of Likert items can be

FIGURE 1 Track Bar Examples



Sources:
MSDN, <http://msdn.microsoft.com/library/default.asp?url=/library/en-us/shellcc/platform/commctls/trackbar/trackbar.asp>

DevX, <http://archive.devx.com/dhtml/articles/nm061102/Hand.html>

combined to form indexes. However, there is a strong caveat to this approach: Most researchers insist such combinations of scales pass the Cronbach's alpha or the Kappa test of intercorrelation and validity.

Also, the combination of scales to form an interval level index assumes this combination forms an underlying characteristic or variable.

Alternative Continuous Measures for Scales

Alternatives to using a formal Likert scale can be the use of a continuous line or track bar. For pain measurement, a 100 mm line can be used on a paper survey to measure from worst ever to best ever, yielding a continuous interval measure.

In the advent of many online surveys, this can be done with track bars similar to those illustrated in Figure 1. The respondents here can calibrate their responses to continuous intervals that can be captured by survey software as continuous values.

Conclusion

Your initial analysis of Likert scalar data should not involve parametric statistics but should rely on the ordinal nature of the data. While Likert scale variables usually represent an underlying continuous measure, analysis of individual items should use parametric procedures only as a pilot analysis.

Combining Likert scales into indexes adds values and variability to the data. If the assumptions of normality are met, analysis with parametric procedure can be followed. Finally, converting a five or seven category instrument to a continuous variable is possible with a calibrated line or track bar.

REFERENCES

1. Gideon Vigderhous, "The Level of Measurement and 'Permissible' Statistical Analysis in Social Research," *Pacific Sociological Review*, Vol. 20, No. 1, 1977, pp. 61-72.
2. Ulf Jakobsson, "Statistical Presentation and Analysis of Ordinal Data in Nursing Research," *Scandinavian Journal of Caring Sciences*, Vol. 18, 2004, pp. 437-440.
3. Rensis Likert, "A Technique for the Measurement of Attitudes," *Archives of Psychology*, 1932, Vol. 140, No. 55.
4. Jum C. Nunnally, *Psychometric Theory*, McGraw Hill, 1978.
5. Dennis L. Clasen and Thomas J. Dormody, "Analyzing Data Measured by Individual Likert-Type Items," *Journal of Agricultural Education*, Vol. 35, No. 4, 1994.

BIBLIOGRAPHY

- Jacoby, Jacob, and Michael S. Matell, "Three-Point Likert Scales Are Good Enough," *Journal of Marketing Research*, Vol. 8, No. 4, 1971, pp. 495-500.
- Jamieson, Susan, "Likert Scales: How to (Ab)use Them," *Medical Education*, Vol. 38, No. 12), 2004, pp. 1,217-1,218.

I. ELAINE ALLEN is an associate professor of statistics and entrepreneurship at Babson College in Babson Park, MA. She has a doctorate in statistics from Cornell University in Ithaca, NY. Allen is a senior member of ASQ.

CHRISTOPHER A. SEAMAN is a doctoral student in mathematics at the Graduate Center of City University of New York.

Please comment

If you would like to comment on this article, please post your remarks on the *Quality Progress* Discussion Board at www.asq.org, or e-mail them to editor@asq.org.