LIKERT SCALE DATA : USE OF DESCRIPTIVE STATISTICS

Jiban Khadka[*]

Abstracts

In this article, the intent is to critique upon some descriptive statistics, especially, mean and standard deviation that have been used to descriptively analyze a Likert Scale data, and to specify and highlight some statistical measures that can be used as alternative measures. As the Likert scale data is, basically, not an interval or a ratio scale data, the statistical assumptions are violated when a social science researcher uses the mean and standard deviation to describe characteristics of the data. Avoiding the misuse of mean and standard deviation in social science research, the researcher can use other non-parametric measures like frequency, percentage, median, mode, range etc. that accounts the ranks or frequency of the scale values. Especially, the frequency percentage is highlighted to use as an effective statistical descriptor in social science research that provides the results about the degree of attitude and its intensity towards any phenomena measured by Likert scale.

Key Words: Likert Scale, Ordinal Scale, Descriptive Statistics, Frequency, Percentage,

[*] **Ph D Scholar, Kathmandu University, Nepal**

Background of the Study

Likert scale or Likert type scale is popular tool of data collection in survey research. It is used to measure the attitude or opinions of individuals towards a phenomena, a subject or an event (Cohen, Manion & Morrison, 2007). A researcher describes a group characteristics based on the data collected from this tool. Despite the increasing popularity of this tool of data collection, the statistics of data analysis has become a matter of controversy for a very long time, and in addition, the major issues are more concerning the use of parametric or non-parametric tests since the date back to 1930s (Norman, 2010). Despite the debates in between the use of parametric and non-parametric statistics, most of researchers are interested to use the parametric measures (Bayage & Lekena, 2010). But, the debates have been little focus on descriptive statistics although the inferential statistics deals with descriptive statistics, and is computed on this basis. For example, correlation, linear regression, t-test and many others are conducted based on the arithmetic mean. Similarly, rank correlation is on the ranks of scale or data, ANOVA on variances and so on. The arguments between two groups of analysts has created confusion on the analysis of Likert scale data. The controversial issues have created dilemma to the novice researchers to use the statistical measures not only in inferential statistics even in descriptive statistics like mean, standard deviation, median, mode etc. In this article, the intent is to critique upon some descriptive statistics, especially, mean and standard deviation that have been used to descriptively analyze a Likert Scale data, and to specify and highlight some statistical measures that can be used as alternative measures. The proposed statistics are expected to preserve the quality of data keeping aside the matter of controversy.

A Quarterly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage, India as well as in Cabell's Directories of Publishing Opportunities, U.S.A.

International Journal of Research in Social Sciences
http://www.ijmra.us

14

Likert Scale and Descriptive Statistics

The Likert scale is constructed based on the principle that the attitude or opinions of an individual can be measured in some grading measuring scales in which respondents express their opinions to favorableness or unfavourableness towards a phenomena (Best & Kahn, 2007). In Likert scale, several statements are administered to measure a variable in which the respondents rate the statements at a certain degree of positive or negative direction like agreement or disagreement. The scale of measurement is categorized by the options in various ways, and commonly, the scales are of five or more than five point Likert scales (Winter & Dodou, 2012). But, almost researchers do not mention the reasons that why the number of scales: five , seven or any other points were chosen (Dolnicar, Grun, Leisch & Rossiter, 2011). It may be assumed that the Likert scale with more number of scales will be closer to specify the information, and they need to contain an exhaustive and discrete range of possible responses that the respondents wish to respond (Cohen, Manion & Morrison, 2007).

In Likert scale, the five point scale, for example, is categorized as strongly agree, agree, undecided, disagree and strongly disagree or the order may be in reverse order like from strongly disagree to strongly agree. In seven point scale, the scales are strongly agree, agree, somewhat agree, undecided, somewhat disagree, disagree and strongly disagree. In attitude scale, there is no legitimation in the response on Likert item that is highly influenced by natural or human factors (Cohen et al., 2007). The matter of studies, in social science, are basically in qualitative nature, and a researcher tries to quantify these qualitative information into quantifiable data such as strongly disagree = 1, disagree= 2, undecided = 3, agree = 4, and strongly agree = 5. In the level of measurement scale, there is not very big debate that the scale is ordinal scale as the

A Quarterly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage, India as well as in Cabell's Directories of Publishing Opportunities, U.S.A.

International Journal of Research in Social Sciences
http://www.ijmra.us

15

scales are used to rank the respondent's attitude towards a subject. The scales rated by the respondents are used for the analysis of data by descriptive or inferential statistics.

In a research, as aforementioned, descriptive statistics are used to describe the characteristics of quantitative data (Gupta, 2007) collected for a group or an experiment. Further, states that the description can be presented in graph, table or as a single value. They describe and present the data in various statistics like frequency, percentage, ratio, maximum, minimum value, mean, median, mode, range, standard deviation, variance, skewness or kurtosis (Cohen et al., 2007). Mean, median and mode are used as to measure central tendency whereas range, quartile deviation, standard deviation, skewness, kurtosis, in absolute or relative form, can be used to measure the dispersion or distribution of data. Mean and standard deviation are used as parametric measures as they used face value (coded value of Likert scale) whereas median, mode, range or quartile deviation are used as non-parametric measures introduced in late 1960s and 1970s (Bayaga & Lekena, 2010) that concerns with positional value. The major issue on the statistics begin from the use of descriptive statistics like mean or standard deviation.

In this paper, the author illuminates the use of descriptive statistics with the assumption of having ordinal nature of Likert scale. Despite the Likert scale as a summated scale that tends to improve the linear and interval scale properties (Perla & Carifio, 2007), no one has claimed with the evidences that the 'scale is interval scale' (Creswell, 2013, p. 167). If a researcher uses mean and standard deviation to describe a mass of Likert scale data, they are in favour of parametric measures. In parametric measures, the values of Likert scale data are used whereas ranks of the values or frequencies are mostly used in non-parametric measures like frequency, median, mode, range and so on. As other assumptions remains the same, the parametric statistics are conducted to analyze the interval or ratio scale data whereas non-parametric statistics to nominal or ordinal

scale data (Creswell, 2013). When the statistics are used to analyze the Likert scale data without ensuring the level of scale that may lead to wrong conclusion. Below the use or misuse of mean and standard deviation are discussed.

Descriptive Statistics: (Mis)use of Mean and Standard Deviation in Likert Scale?

Several researchers treat Likert scale data at the interval measurement scale tests (De Winter and Dodou, 2012) and the scale is created by calculating a composite score (sum or mean) from many items. The descriptive statistics recommended for interval scale items include the mean (mostly arithmetic mean) for central tendency and standard deviations for variability(Boone & Boone, 2012) with the assumption of Likert scale as interval scale. In interval scale, the intervals are assumed as equal distance, but it is logically not possible to justify as equal intervals between two scales. Although the intervals may be any real numbers, it can not be claimed as equal intervals in Likert scale. For example, 1, 2, 3, 4 and 5 be five scales that represent strongly disagree, agree, undecided, agree and strongly agree respectively . In which, the intervals in 2-1, 3-2, 4-3 or 5-4 are real numbers and seem numerically equal. But, the numbers 1, 2, 3, 4 and 5 are used just for coding. The result is not true that disagree - strongly agree equals to undecided-agree (Cohen et al., 2007) as the difference is qualitative matter. This is the fundamental issue of Likert scale. The recommendation for mean and standard deviation for Likert scale is not appropriate (Jamieson, 2004) without justification of Likert scale or its sum as interval scale. This also contradicts to the version that combined score of Likert scale tends to interval scale, and can be calculated as the exact difference between scales. An example that includes two types of data are presented in table 1.1 to make this concept more clear.

The example presented in Table 1.1 shows the data analysis to investigate the motivational level of school principals as perceived by the 100 school teachers. The results is

yielded by the use of SPSS version 16. The data are brought for the discussion in which age detail and their responses towards the five items in five point Likert scale that would measure the motivational level of their principals are as follows:

Table 1.1 Frequency of age of respondents and their responses on Likert Scale

Age of Respondents	No. of Respondents	Likert Scale Response		No. of Responses
25	29	Strongly Disagree	(1)	25
26	41	Disagree	(2)	18
27	13	Undecided	(3)	87
28	9	Agree	(4)	198
29	8	Strongly agree	(5)	172
	Mean = 26.26	Mean = 3.94		

In the study, there were 100 cases with the age of 25, 26, 27, 28 and 29 years. In the measurement of ages, intervals are equal like 26-25 (= 1 year) =27-26 (1 year) = 28-27 (1 year), 29-28 (=1 year), and the intervals between any two consecutive ages are equal or an interval length always scalar times the another interval length in any group of respondents. In the analysis of Likert scale response, the intervals are equal like 2-1 = 3-2= 4-3 =5-4 (difference is 1). In the measurement of age, by nature, it is numerical data, and the standard units can be used to measure age in years, months, days, hours, minutes or seconds. It can be reduced into the small units as we wish or as big units as we wish. But, in Liker scale response, there is no standard unit to measure the difference between two scales and obviously can not be factorized into small numbers as we wish. One confusing matter might be that what are 1, 2, 3, 4 and 5 used in Likert scale. As aforementioned, the numbers are numerical symbols that are used to rank the the attitude of respondents, and it doesn't mean that 1 is standard unit to measure 'strongly disagree' likewise the measurement of age. The human behaviours can be compared in terms of more or

A Quarterly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage, India as well as in Cabell's Directories of Publishing Opportunities, U.S.A.

International Journal of Research in Social Sciences
http://www.ijmra.us

18

less, better or worse, but not possible to compare as exact science like greater by 'this much quantity'. These evidences limit that the Likert scales are ordinal scales (greater than form) but not the interval scale (Boone & Boone, 2012).

However, if we follow the assumption of being equal intervals of measurement scale for the use of mean, it best fits for the Likert scale analysis, and the mean is 3.94. When we interpret the mean 3.94, we say, in average, the respondents more than undecided level and closer to agree. The meaning of .94 is not clear that actually what does .94 represent? Because, .94 is not the unit of Likert scale measurement and which point represents .94 on the Likert scale or how much part of the difference between undecided and agree can be represented by .94, and similar arguments are also raised by Cohen et al. (2007). In contrast, the mean of age is 26.26 which means that the average age of respondents is more than 26 years. If we need the meaning of .26, it represents 94 days, 21 hours and 36 minutes as the measurement scale of age is continuous. This justifies that the assumption of interval or ratio scale is necessary condition for mean to produce meaningful results.

In computing mean or any other similar statistics, several scores are combined both directions from agreement to disagreement but that fails to explain the intensity of response well (Albaum, 1997, p. 334). Although the human study is also based on natural science with the assumption that the human behaviour are measurable, empirical and replicable and govern by some laws or theory (Pant, 2012),in social science, as stated by Gaito (1980), in real sense, the constructs even variables are not numerical phenomena (Norman, 2010). It means that human behaviour can not be measured in continuous scale rather they can be categorized into ranks: more or less. The result based on inappropriate measure like mean is obviously not acceptable. In similar manner, standard deviation gives information more or less about the dispersion from

the mean. When mean is not suitable to measure central tendency, the dispersion measured by standard deviation is not good or accurate descriptor in Likert scale. As aforementioned, avoiding the misuse of mean and standard deviation in social science, the researcher can use other non-parametric measures that accounts the ranks of the scale values like frequency, percentage, median, mode, range etc. Then, the frequency percentage can be used as a more alternative measure in social science research is discussed as follows:

Descriptive Statistics: How to use frequency percentage in Likert Scale Data?

The subject of social science is basically qualitative in nature and the responses on each scale can have important role to explain the phenomena as 'only crude data can result the crude interpretation' (Cohen et al., 2007). In Likert scale, only 25 respondents (from table 1.1) were strongly disagreed which is 5% of the total responses that is not highlighted while using the mean. When we analyze the Likert scale data from positivist perspective, the mean may be enough to describe but the mean will not be effective to address the research issues in social science as it ignores the intensity and hides the responses around at extreme level (Sullivan & Artino, 2013; Albaum , 1997). Despite the alternative measures that better portrays the picture from the analysis of Likert scale, it is surprising that why the scholars in favour of parametric measures like mean or standard deviation rather than frequency, percentage, median, mode, range, quartile deviation or its coefficients that seem more appropriate to observe the phenomena logically. Further, as stated by Cohen et al. (2007), it is not fair to treat non-parametric data as parametric violating the assumptions of the statistics. The frequency or percentage of responses on each scale can be calculated without violating the statistical assumption, and the statistics can provide understandable and more explicit information rather than mean and standard deviation.

Although the quantitative research is a scientific research that deals with the single reality, the phenomena in social science is not like physical object or event rather it is inherently qualitative and needs to be treated as qualitative phenomena without distorting its crude nature. The analysis of each scale based on non-parametric measures like frequency percentage, range or similar measures can have meaningful contribution to address the research issues that retains, to some extent, the strength of qualitative information too. Because, several items are used to measure a social construct in objective way and each item or each respondent has contribution to enrich the quality of research although all the items or respondents may not be possible to be analyzed individually due to several constraints like cost, time and effort. However, this process of analysis helps a researcher to understand more objectively and subjectively as well. Further, it enables the researcher to identify the strong or weak feeling of respondents.

We rarely find the research studies where the researchers like Ivankova & Stick (2007) as cited in Creswell & Plano Crark (2011) have used frequency or pencentage as descriptive statistics. In author's opinion, the use of frequency percentage can be effective to describe the information drawn from Likert scale. The Likert scale is used to measure attitude towards two aspects of a phenomena: degree and intensity. In measurement of degree of response, it is either positive or negative. In positive direction, there might be any options on the scale like agree, acceptable, appropriate, true, desirable or satisfied, and in negative direction, the options may be disagree, unacceptable, inappropriate, untrue, undesirable or dissatisfied. In frequency measuring scale, if the statement is affirmative, the terms: almost every time, every time and so on can measure the positive direction and in negative direction, not al all, never, rarely and so on can be used. The another part of scale is the mid-point that has important role to the result of Likert scale data. Several terms are used that are used as the mid-term scale like undecided, neither...

nor...., neutral, unsure and many others. An analyst has to be attentive to these state of responses: positive, negative and mid-point. From these three scales, direction of responses on Likert scale can be portrayed. With the assumption of Likert scale as ordinal scale, 'degree of agreement' can be used to analyze the data (Creswell, 2013). The simplest way of analyzing the data is to categorize the five or more scales into three categories (Best & Kahn, 2007): agree, undecided and disagree. In the separation of categories, all the scales can be merged into three categories like agree, disagree and undecided or positive, negative and mid-point responses. In the following example, strongly agree and agree are merged into agree, strongly disagree and disagree are into disagree and the undecided response remains the same. Then, the result of data analysis is as follows:

Table 1.2 Frequency Percentage of Responses on Five Point Likert Scale

Likert Scale		Frequency Percent	Merged Scale	Merged Percentage
Strongly Disagree	(1)	5%	Disagree	8.6%
Disagree	(2)	3.6%		
Undecided	(3)	17.4%	Undecided	17.4%
Agree	(4)	39.6%	Agree	74%
Strongly agree	(5)	34.4%		

Overall: Mean = 3.94, Median = 4, Mode = 4, s.d. = 1.054, S_k = -1.110

From table 1.2, the table shows that merged percentage shows that the percentage (74%) on agree is far greater than disagree (8.6%) and the the intensity (39.6%) is the highest on agree (rating scale =4). This result is consistent with median and mode as they (median or mode) lie in the first or second rating positions of intensity: agree or strongly agree. It is also easily to show that the distribution is more skewed to left due to the mode is higher at 4 or might be closer to the value 5 although it is of less concern in non-parametric measures. Out of three categories of responses, 17.4% of the respondents are at the level of undecided that the reluctant rating is not possible to identify by the use of mean and standard deviation. Although the mid value does not

explain the charatceristics of a mass of data, the mid-value has effect on some descriptive statistics (Sullivan & Artino, 2013). In the example, how mean, median or mode are affected by mid-scale is presented on the following table 1.3:

Table 1.3 Comparison of Mean, Median and Mode between the groups of inclusion and exclusion of Mid-value Response

Statistics	Item1	Item2	Item3	Item4	Item5	Mid Scale
Mean	4.0	4.0	3.6	4.0	3.9	Included
Mean	4.2	4.2	3.9	4.1	4.1	Excluded
Median	4	4	4	4	4	Included
Median	4	4	4	4	4	Excluded
Mode	5	4	4	4	5	Included
Mode	5	4	4	4	5	Excluded

Table 1.3 shows that the inclusion or exclusion of mid-scale response has no or little numerical effect on median and mode . But it affects the mean value despite insignificant difference between the means of the groups of inclusion and exclusion of mid-scale response at 5% level of significance as the p $(= .767) > \alpha = .05$. The numerical difference in means can affect on the descriptive analysis and creates the confusion to interpret the results (Sullivan & Artino, 2013).

Conclusion

The Likert scale has been used conventionally to measure the attitude of people towards the any subject of study. The attitude is not inherently quantitative phenomena rather it is qualitative and then subjective. The less debated issue is that it can be measured in ordinal scale but not in continuous scale by Likert scale. It is used to measure as to positive or negative direction, and about the intensity of attitude as well. As the Likert sale data, basically, is not interval or ratio

A Quarterly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage, India as well as in Cabell's Directories of Publishing Opportunities, U.S.A.
International Journal of Research in Social Sciences
http://www.ijmra.us

23

scale data, the statistical assumptions are violated when a social science researcher uses the mean and standard deviation to describe characteristics of the data. These measures (mean and standard deviations) are more or less fluctuated by the inclusion or exclusion of some responses that may hide the original distribution of the data. As the best descriptors, the frequency or percentage on each scale of Likert Scale, and on three categories: positive, negative and mid-scale can be used which is the theoretical expectation of Likert scale. They provide the information about both degree of attitude and its intensity towards the subject. If the single value as central tendency is the indicator of the study, the median or mode can be used as they are less affected by the fluctuation of Likert scale response. The use of frequency, percentage, median and mode can preserves the ordinal nature of Likert scale data and enables the researchers to describe attitude towards any phenomena.

Reference

Albaum, G. (1997). The Likert scale revisited: an alternate version. *Journal of the market research society,* 39 (2).

Bayaga & Lekena (2010). Application of non-parametric analysis technique amongst postgraduate education research: A survey of South African Universities. *The journal of international social research,* 3(11).

Best, J. W. & Kahn, J. V. (2007). *Research in education* (9th ed.). New Delhi: Prentice-Hall.

Boone, N. H. & Boone, D. A. (2012). Analyzing Likert data. *Journal of extension*, 50 (2). ISSN 1077-5315.

Cohen, L., Manion, L. & Morrison, K. (2007). *Research methods in education* (6th ed.). USA: Routledge Publication.

A Quarterly Double-Blind Peer Reviewed Refereed Open Access International e-Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage, India as well as in Cabell's Directories of Publishing Opportunities, U.S.A.
International Journal of Research in Social Sciences
http://www.ijmra.us

24

Creswell, J. W. (2013). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (4th ed.). Delhi: PHI Learning Private Limited.

De Winter and Dodou (2012). Five-Point Likert Items: t test versus Mann-Whitney-Wilcoxon. *Practical Assessment, Research & Evaluation, Vol 15, No 11,* ISSN 1531-7714.

Dolnicar, S., Grun, B. Leisch, F. & Rossiter, J. (2011). Three good reasons not to use five and seven point Likert items. This conference paper publised in CAUTHE 2011 : 21st CAUTHE National Conference, Adelaide, Australia, 8-11 February 2011.

Gupta, S. C. (2007). *Fundamentals of statistics* (6th ed.). Mumbai: Himalaya Publishing House.

Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical education*, 38, 1212–1218. doi: 10.1111/j.1365-2929.2004.02012.x.

Norman, G. (2010). Likert scales, levels of measurement and the ''laws'' of statistics. Adv in Health Sci Educ. Springer Science+Business Media B.V.

Perla , R. J. & Carifio, J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about likert scales and likert response formats and their antidotes**.** *Journal of Social Sciences* 3 (3): 106-116, 2007**.** ISSN 1549-3652.

Sullivan, G. M. & Artino, A. R. (Ed.) (2013). Analyzing and interpreting data from likert-type scales (Editorial). *Journal of Graduate Medical Education*. DOI: http://dx.doi.org/10.4300/JGME-5-4-18.