

Extending the Consensus Measure: Analyzing Ordinal Data With Respect to Extrema

Jennifer M. Tastle
Jtastle1@ithaca.edu
Dept of Accounting

William J. Tastle
tastle@ithaca.edu
School of Business

Ithaca College
Ithaca, New York 14850, USA

Abstract

The existence of an ordinal measure already exists and is well justified, but a modest extension of the consensus formula permits Likert scale data to be assessed with respect to a predetermined extreme value, and the results used for comparisons and trajectories. The new measure, called the strength of consensus measure, is a modification of both the Shannon entropy, an equation common to the foundation of information theory, and the standard consensus measure.

Keywords: strength of consensus, agreement, consensus, ordinal measure, entropy

1. INTRODUCTION

Classroom activities involve the assessment of students by means of different measures, typically by using the Likert scale. It is common for students to be asked to critique their peers or work groups, responding to questions by filling in the appropriate bubble sheet entry, i.e., strongly agree (SA), agree (A), neutral (N), disagree (D), and strongly disagree (SD). There is considerable research concerning methods of motivating teams in agreeing on identifying business problems, however there is little research in the formulation of mathematical measures that can guide teams effectively (Tastle, Wierman, and Dumdum, 2005).

Utilizing the Shannon (1948) entropy:

$$H(X) = -\sum_{i=1}^n p(x_i) \log_2 p(x_i) \quad (1)$$

where X is the set of n categories under investigation, and p_i is the probability of each x_i , a new measure of dispersion has been developed (Wierman and Tastle, 2005). A comprehensive description of the entropy measure and its properties can be found in Klir and Wierman (1997). However, this entropy measure does not rank-order the x_i values. Thus, every permutation of values yields the same entropy value, and that is the short-fall of using the Shannon entropy measure for assessing the Likert scale.

The new measure (see equation 2) provides a means by which ordinal scale data can be assessed with respect to its dispersion around a central mean. The measure can be

extended, however, by fixing the mean value to a predetermined focal point and then calculating the measure. This feature is illustrated in this paper and confined to the extreme values in the Likert scale, herein referred to as the extrema of a set or ordered categories.

We describe the different kinds of measurement scales, provide a brief description of the new measure, and then extend the measure to permit an analysis of one kind of ordinal data measure, specifically that of a "weighted" Likert scale.

2. THE MEASURE OF CONSENSUS

Consensus is a term used to describe a group's shared feelings toward a particular issue. For purposes of this paper, we assume a simulated set of five groups of students, each group composed of four students. We seek to determine, by means of a Likert scale measure, how the individual group members feel about the success within their particular team. We then want to order the teams based on the success, and of their activities. Currently, such an endeavor would be difficult at best, for the measures available are inappropriate and incomplete for the most part. We approach this task by fixing the mean value of the team members Likert scale evaluations to a predetermined focal point and then calculate the consensus.

The consensus measure is defined as:

$$Cns(X) = 1 + \sum_{i=1}^n p_i \log_2 \left(1 - \frac{|X_i - \mu_X|}{d_X} \right) \quad (2)$$

where X is the Likert scale, p_i is the probability of each X , d_X is the width of X , X_i is the particular Likert attribute, and μ_X is the mean of X (Wierman and Tastle, 2005). The mean, μ_X , is the expected value,

$E(X) = \sum_{i=1}^n p_i X_i$. However, the mean value can "float" the entire length of the Likert scale, from strongly agree to strongly disagree depending on the values of p_i . We seek to assess each group of students based on the individual student's perception of the overall quality of their group as described by the set of questions each student must answer. Thus, for each group of four students there is an individual Likert scale for each question. It is possible that the

average perception for the students to be any value between SA and SD, the extreme points on a Likert scale, and each group can have a distinctly different mean value. This complexity essentially prevents us from comparing the groups. We need to calculate a measure based on the same focal point, i.e., a generally accepted central value from which the consensus can be measured. We shall arbitrarily decide to use *strongly agree* as the focal point, and we insist that each Likert scale question be written in a positive tone. We expect that each student in each group would ideally like to strongly agree with each question.

Thus, given a statement "The team worked well together," it is most desirable to have the entire team membership check the *strongly agree* bubble. If that does occur, then all members are in 100% consensus on that particular item. Realistically, the team members can select any combination of Likert scale values. The illustration in the next section shows the interactions of this measure and how different teams can be compared.

3. THE STRENGTH OF CONSENSUS

The variation of the current consensus measure (Tastle and Wierman, 2005; Wierman and Tastle, 2005) is attained by increasing the system width, d_X , to $2(d_X)$, and fixing μ_X to 1.

$$sCns(X) = 1 + \sum_{i=1}^n p_i \log_2 \left(1 - \frac{|X_i - 1|}{2d_X} \right) \quad (3)$$

The resulting equation permits us to calculate the strength of consensus, $sCns$. The original consensus measure failed when the focus was either extreme.

Illustration of the Strength of Consensus Measure

By assigning the mean value to a focal point such as *strongly agree*, the consensus value is focused with respect to that point. Thus, instead of a meandering weighted mean value as currently exists, the original consensus measure, the focal point is required to always be *strongly agree* and assigned a value of 1. Thus, SA = 1, A = 2, N = 3, D = 4, and SD = 5. If the weighted mean was calculated, as in the regular consensus measure, the value would be

contained in the range 1 to 5 with very little opportunity being given to either extreme value. As students record their perceptions of the team activity, the data can be tallied to determine an overall team score. Essentially, using the consensus measure, the SA, A, etc. scores are replaced with a single real number that captures the meaning of the Likert values. Thus, if the majority of the scores center around neutral or disagree, the focal point (SA) will be a greater distance from the category values and the resulting strength of consensus will be less (closer to 0). However, if the majority of students are in strong agreement or agree, then the strength of consensus will be close to 1.

To illustrate, use our previously created fictitious classroom of five groups of four students each. Projects are regularly assigned to these groups throughout a semester; we assume that group membership does not change. After the completion of each assignment, students are given a consistent questionnaire containing the four evaluation criteria listed below. Each student answers all the questions by filling a bubble sheet using a Likert scale format (strongly agree, agree, neutral, disagree, strongly disagree). The criteria are:

1. The team worked very well together.
2. All individual members performed their tasks to an acceptable high standard.
3. The team collectively accomplished the task.
4. All task requirements were completed on time.

After each questionnaire is given, the data from each student is tallied to create a set of values for the respective group. Table 1 shows the result of the tally and the calculation of the strength of consensus:

Table 1 Tally of activity 1 data

| Group | SA | A | N | D | SD | sCns |
|-------|----|---|---|---|----|------|
| 1 | 0 | 7 | 3 | 2 | 4 | 50% |
| 2 | 1 | 7 | 3 | 4 | 1 | 61% |
| 3 | 0 | 7 | 3 | 4 | 2 | 54% |
| 4 | 2 | 6 | 2 | 2 | 4 | 54% |
| 5 | 2 | 9 | 3 | 2 | 0 | 73% |

Illustration of calculation

Utilizing equation 3, Group 1 will be used to illustrate how the strength of consensus was calculated to achieve a value of 50%.

Recall that the strength of consensus is based on the presumption that the mean value is "Strongly Agree," and hence has a numerical value of 1 (SA = 1, A = 2, etc.). Thus, μ_x will always equal 1. Equation 3 of Table 1 is calculated as though it was a normal consensus measure, except that the mean is static at 1.

For $i = 1$ (that is to say, the SA value), $p_i = 0$.

For $i = 2$, the presence of a non-zero value permits a calculation:

$$= .44 \log_2 \left(1 - \frac{|2-1|}{(2)(4)} \right) = .44 \log_2 (.875) \\ = -0.085$$

For $i = 3$:

$$= .19 \log_2 \left(1 - \frac{|3-1|}{(2)(4)} \right) = .19 \log_2 (.75) \\ = -0.079$$

For $i = 4$:

$$= .13 \log_2 \left(1 - \frac{|4-1|}{(2)(4)} \right) = .13 \log_2 (.625) \\ = -0.089$$

For $i = 5$:

$$= .25 \log_2 \left(1 - \frac{|5-1|}{(2)(4)} \right) = .25 \log_2 (.5) \\ = -0.25$$

We sum these results and apply equation 3 to get the strength of consensus percentage: $sCns(X) = 1 + (0 - 0.085 - 0.079 - 0.089 - 0.25) = 1 + (-0.503) = 0.497 = 50\%$

Examining the individual rows of data (Table 1) we observe that the data are scattered such that it is not immediately known how to best rank the groups. Observing groups 3 and 4, we note that group 3 has no strongly agree values while group 4 has two. We

also note that group 4 has four strongly disagree votes while group 3 only has two. With the addition of the strength of consensus column, it appears that both groups have rated their performance level at 54%, in spite of their varying data. Group 3 is more centered around agree and neutral while group 4 is spread out across the scale. One may think that because group 4 has two strongly agrees that that group is the "better" group when in reality, it is no different than group 3. A second questionnaire is given after another group project and the results are shown in Table 2:

Table 2 Tally of activity 2 data

| Group | SA | A | N | D | SD | sCns |
|-------|----|---|---|---|----|------|
| 1 | 3 | 8 | 0 | 4 | 1 | 67% |
| 2 | 4 | 5 | 0 | 4 | 3 | 58% |
| 3 | 4 | 5 | 4 | 0 | 3 | 65% |
| 4 | 4 | 5 | 2 | 0 | 5 | 58% |
| 5 | 2 | 7 | 4 | 3 | 0 | 68% |

Again, although the data are scattered, the strength of consensus among some of the groups is similar. Groups 2 and 4, although very different data, both have 58% consensus on their group's performance level. One can compare the strength of consensus from both tables and determine whether a group's consensus on their performance increased or decreased. For group 5, although only two group members strongly agreed that their group performed well, it has the highest consensus level. The more each member within one group can agree that they performed well, the higher the consensus.

Combining the ranking from these two examples (Table 3) permits a side-by-side comparison of Likert scale results.

| Table 3 Combined data | | | |
|------------------------------|-------|-------------------|-------|
| Activity 1 | | Activity 2 | |
| sCns | Group | sCns | Group |
| 73% | 5 | 68% | 5 |
| 61% | 2 | 67% | 1 |
| 54% | 3 | 65% | 3 |
| 54% | 4 | 58% | 2 |
| 50% | 1 | 58% | 4 |

It is noted that team 5 is clearly in the lead after two activities, and that teams 1 and 2

have switched places. Team 1 has greatly improved, but team 2 appears to be in trouble. The instructor can now direct his/her attention to this team.

Rather than treating the data from tables 1 and 2 as individual projects, we combine the data from these two projects together to identify an overall level of performance. The calculation of sCns is taken over the total project set. This permits the instructor to capture the overall performance of the groups. The analysis of this aggregate data is shown in Table 4. It is easy to see that group 5 is in the lead with the highest ranking. Group 3 is ranked in second position, a fact that was not identifiable from table 2.

Table 4 Combined tally of table 1 and table 2 data

| Group | SA | A | N | D | SD | sCns |
|-------|----|----|---|---|----|------|
| 1 | 3 | 15 | 3 | 6 | 5 | 59% |
| 2 | 5 | 12 | 3 | 8 | 4 | 59% |
| 3 | 4 | 12 | 7 | 4 | 5 | 60% |
| 4 | 6 | 11 | 4 | 2 | 9 | 56% |
| 5 | 4 | 16 | 7 | 5 | 0 | 71% |

Adding each group data from Table 1 and Table 2 shows us that although group 4 has the greatest number of students who strongly agree, they have the lowest strength of consensus factor. This enforces the strength of consensus measure in that there needs to be a scales in which data can be measured.

4. CONCLUSION

This paper discusses how ordinal scale data can be assessed with respect to its relation to a specific value. The example illustrates the application of a variation in the consensus measure, called the strength of consensus, which permits an ordinal measure to be represented by a real number and compared to other ordinal scales.

Future research to determine if the number of categories affects the ranking of values should be undertaken. This characteristic is called granulation, and while some work has been done, it did not have the benefit of this new measure. Lastly, work should be done to see if the strength of consensus measure

is sufficient, by itself, to adequately substitute for the weighted mean and standard consensus measure.

5. REFERENCES

- Klir, George J. and Mark J. Wierman, (1997). Lecture Notes in Fuzzy Mathematics and Computer Science: Uncertainty-Based Information Elements of Generalized Information Theory. Center for Research in Fuzzy Mathematics and Computer Science, Creighton University, Omaha, Nebraska.
- Tastle, William J and Mark J Wierman (2005b), "Consensus and Dissension: A New measure of Agreement." North American Fuzzy Information Systems Processing Conference (NAFIPS), Ann Arbor, MI.
- Tastle, William J., Mark J Wierman and U. Rex Dum Dum (2005), "Ranking Ordinal Scales Using The Consensus Measure: Analyzing Survey Data." IACIS 2005.
- Wierman, Mark J. and William J. Tastle (2005), "Consensus and Dissension: Theory and Properties." North American Fuzzy Information Systems Processing Conference (NAFIPS), Ann Arbor, MI.