Bagadi, R. (2017). Total Intra Similarity And Dissimilarity Measure For The Values Taken By A Parameter Of Concern. {Version 2}. ISSN 1751-3030. *PHILICA.COM Article number 1153*. http://philica.com/display_article.php?article_id=1153

Total Intra Similarity And Dissimilarity Measure For The Values Taken By A Parameter Of Concern. {Version 2}. ISSN 1751-3030

Author:

Ramesh Chandra Bagadi
Data Scientist
INSOFE (International School Of Engineering),
Hyderabad, India.
rameshcbagadi@uwalumni.com
+91 9440032711

Research Manuscript

Abstract

In this research investigation, the author has detailed a novel method of finding the '*Total Intra Similarity And Dissimilarity Measure For The Values Taken By A Parameter Of Concern*'. The advantage of such a measure is that using this measure we can clearly distinguish the contribution of Intra aspect variation and Inter aspect variation when both are bound to occur in a given phenomenon of concern. This measure provides the same advantages as that provided by the popular F-Statistic measure.

Theory

Given any Sequence of the kind,
$S=\{y_1, y_2, y_{3,} ...., y_{n-1}, y_n\}$ which represent the (ordered) values taken by a Parameter of concern. We first consider the Cartesian Cross Product of S with itself, i.e. $R=S\times S=\underset{i=1}{\overset{n}{C}}\underset{i=1}{\overset{n}{C}}\{y_i, y_j\}$ where C denotes a Collection. We now consider elements of the kind $\{y_i, y_j\}$ and $\{y_j, y_i\}$ as same and therefore will consider only one among them in the sum terms shown below in the next line, which appears as

$$LB=\underset{i=1}{\overset{n}{C}}\underset{i=1}{\overset{n}{C}}\{y_i, y_j\}-\underset{j:j>i}{\overset{n}{C}}\ \underset{i=1}{\overset{n}{C}}\{y_i, y_j\} \quad \text{Equation 1}$$

We now consider the smaller of each of the 2 tuple in the above set LB and add these values to give us

$$SLB=\sum\left\{Smaller\left\{\underset{i=1}{\overset{n}{C}}\underset{i=1}{\overset{n}{C}}\{y_i, y_j\}-\underset{j:j>i}{\overset{n}{C}}\ \underset{i=1}{\overset{n}{C}}\{y_i, y_j\}\right\}\right\} \quad \text{Equation 2}$$

For the terms of the kind $\{y_i, y_i\}$, the Smaller Operator detailed above selects y_i itself. The spirit behind considering the smaller number is that it represents the congruence part of the two numbers. The logic behind removing the $\underset{j:j>i}{\overset{n}{C}}\ \underset{i=1}{\overset{n}{C}}\{y_i, y_j\}$ terms is that when their congruence part is evaluated it is the same as their places juxtaposed counterparts of themselves in $S\times S$.

We now consider the larger of each of the 2 tuple in the above set LB (RHS of Equation 2) and add these values to give us

$$LLB = \sum \left\{ Larger\left\{ \underset{i=1}{\overset{n}{C}} \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} - \underset{j:j>i}{\overset{n}{C}} \ \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} \right\} \right\}$$

Equation 3

Now, we call the Total Intra Similarity

Measure as

$$AD = \frac{SLB}{LLB} = \left\{ \frac{\sum \left\{ Smaller\left\{ \underset{i=1}{\overset{n}{C}} \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} - \underset{j:j>i}{\overset{n}{C}} \ \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} \right\} \right\}}{\sum \left\{ Larger\left\{ \underset{i=1}{\overset{n}{C}} \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} - \underset{j:j>i}{\overset{n}{C}} \ \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} \right\} \right\}} \right\}$$

Equation 4

For the terms of the kind $\{y_i, y_i\}$, the Larger Operator detailed above selects y_i itself. The advantage of such a measure AD is that using this measure we can clearly distinguish the contribution of Intra aspect variation and Inter aspect variation (with respect to their similarity) when both are bound to occur in a given phenomenon of concern. This measure provides the same advantages as that provided by the popular F-Statistic measure.

Furthermore, we can even write the Total Intra Similarity Vector as

$$TISV = \sum_k \left\{ Smaller\left\{ \underset{i=1}{\overset{n}{C}} \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} - \underset{j:j>i}{\overset{n}{C}} \ \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} \right\} \right\} \hat{e}_k$$

Equation 5

where $k = n^2 - \left(\frac{n^2 - n}{2} \right) = \frac{n^2}{2} + \frac{n}{2}$

We can use this Total Intra Similarity Vector to compare the Total Intra Similarity Measure of two Parameters that take the same number of values. Such a comparison can be achieved using any popular type of Inner Product scheme. These two Parameters could be different or could be the same Parameter whose observation is repeated again.

In a similar manner, we can define the Total Intra Dissimilarity Measure as

$$AD^- = \left(\frac{LLB - SLB}{LLB} \right)$$

$$= \left\{ \frac{\sum \left\{ Larger\left\{ \underset{i=1}{\overset{n}{C}} \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} - \underset{j:j>i}{\overset{n}{C}} \ \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} \right\} - Smaller\left\{ \underset{i=1}{\overset{n}{C}} \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} - \underset{j:j>i}{\overset{n}{C}} \ \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} \right\} \right\}}{\sum \left\{ Larger\left\{ \underset{i=1}{\overset{n}{C}} \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} - \underset{j:j>i}{\overset{n}{C}} \ \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} \right\} \right\}} \right\}$$

Equation 6

The spirit behind considering the difference between the larger and smaller number is that it represents the non-congruence part of the two numbers.

Bagadi, R. (2017). Total Intra Similarity And Dissimilarity Measure For The Values Taken By A Parameter Of Concern. {Version 2}. ISSN 1751-3030. *PHILICA.COM Article number 1153*. http://philica.com/display_article.php?article_id=1153

The advantage of such a measure AD^{-} is that using this measure we can clearly distinguish the contribution of Intra aspect variation and Inter aspect variation (with respect to their dissimilarity) when both are bound to occur in a given phenomenon of concern. This measure provides the same advantages as that provided by the popular F-Statistic measure.

Furthermore, we can even write the Total Intra Dissimilarity Similarity Vector as

$$TIDSV = \sum \left\{ \left\{ \begin{array}{l} Larger\left\{ \underset{i=1}{\overset{n}{C}} \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} - \underset{j:j>i}{\overset{n}{C}} \ \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} \right\} \\ - Smaller\left\{ \underset{i=1}{\overset{n}{C}} \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} - \underset{j:j>i}{\overset{n}{C}} \ \underset{i=1}{\overset{n}{C}} \{y_i, y_j\} \right\} \end{array} \right\} \right\} \hat{e}_k$$ where

$$k = n^2 - \left(\frac{n^2 - n}{2} \right) = \frac{n^2}{2} + \frac{n}{2}$$ and the summation is over the k terms.

We can use this Total Intra Dissimilarity Vector to compare the Total Intra Dissimilarity Measure of two Parameters that take the same number of values. Such a comparison can be achieved using any popular type of Inner Product scheme. These two Parameters could be different or could be the same Parameter whose observation is repeated again.

References

1.http://www.philica.com/advancedsearch.php?author=12897
2.http://www.vixra.org/author/ramesh_chandra_bagadi