The Cluster Sensitivity Index: A Basic Measure of Classification Robustness

Willard C. Hom
California Community Colleges

Analysts of institutional performance have occasionally used a peer grouping approach in which they compared institutions only to other institutions with similar characteristics. Because analysts historically have used cluster analysis to define peer groups (i.e., the group of comparable institutions), the author proposes and demonstrates with actual peer grouping data a method for diagnosing the robustness of a peer grouping effort. This method measures the effect of alternative clustering approaches through the calculation of the cluster sensitivity index (CSI). This index, however, is flexible enough to help in other kinds of comparisons that do not apply cluster analysis or do not involve institutions of higher education.

State offices of higher education and policy analysis groups are two entities that often have an interest in evaluating the performance of post-secondary institutions. However, post-secondary institutions present a major challenge to such evaluations because they differ in many substantive ways, such as enrollment size, mission, and budget, to mention a few. One common approach to support a valid evaluation of institutional performance is the comparison of an institution's performance to the performances of its peer institutions (Hurley, 2002; Goan, 2007). The Carnegie classification has gained historical significance, largely because of its function in aiding in institutional comparisons and evaluations (Carnegie, 2007; McCormack & Zhao, 2005; McCormack & Cox, 2003). The National Center for Education Statistics has designed and maintained a fairly inclusive peer group process that has widespread access and usage among institutional researchers (U.S. Department of Education, 2007). California has output a peer grouping of its 109 community colleges since 2007 as part of its legislatively mandated accountability program (Hom, 2008; California Community Colleges, Chancellor's Office, 2008).

Analysts can determine an institution's peer group through a number of different methods. One popular method for peer grouping is cluster analysis (Phipps, Shedd & Merisotis, 2001). For details on this statistical technique, see other references for expert explanations (Aldenderfer & Blashfield, 1984; Lorr, 1987; Everitt, Landau & Leese, 2001; Hair, et al., 2006; Kettenring, 2006). However, cluster analysis, like other classification methods, can assign to an institution different peers as the analyst varies two critical elements in cluster analysis. These two elements are (1) the measure of proximity and (2) the clustering algorithm. The choices of a proximity measure and of clustering algorithm have been shown to produce distinctly different peer groups for certain data sets (Green & Rao, 1969; Funkhouser, 1983; Frank & Green, 1968).

Analysts can encounter a dilemma, in that a particular institution can have one set of peers with one cluster analysis method and a different set of peers with another cluster analysis method. From a technical viewpoint, such varying results are evidence of *method bias* in an analysis (Everitt, et al., 2001; Ketchen & Shook, 1996). From the author's experience, a valid peer group definition would show little evidence of method bias. It would have a *robust definition* of a peer group, meaning that different clustering approaches will produce the same (or nearly the same) set of peers for a specific institution.

If the identification of peers were only an academic exercise, the method bias could be accepted as a common characteristic of statistical analysis, and it could be left at that. However, policy-makers (such as higher education administrators and legislators) may want to make decisions that use performance evaluations that depend upon some type of comparison. For this group, the analyst must address the following question:

> How much confidence can someone place in the use of a cluster analysis to find a peer group for a given institution?

In the following sections, the author proposes and demonstrates one possible answer to the above question. This proposal may benefit many analysts/evaluators because the CSI (a) requires relatively simple calculations, (b) has intuitive appeal, and (c) generalizes to comparisons of classifications that use methods other than cluster analysis.

Related Methods

The CSI tries to fill a gap in the tools for diagnosing classification results. Other tools try to evaluate the validity of cluster groupings that someone has generated (Jain & Dubes, 1988; Xu & Wunsch, 2009). These differ from the CSI because they focus upon the amount of separation between the analyst-identified peer groups or clusters. These validation efforts also tend to focus upon the results of one classification method. Furthermore, the CSI specifically addresses the robustness in classification for a single member of a population undergoing classification. In many decision-making situations, the analyst only has concern about the robustness of a peer group definition in terms of his/her own institution (not the robustness of the clustering in terms of all of the institutions in a population).

Other tools seek to optimize the classification results through algorithms that exploit a comparison of classification results with known groupings. Neural networks, discriminant analysis, classification and regression trees, and genetic algorithms use this approach (Hand, Mannila & Smyth, 2001; Han & Kamber, 2001; Kettenring, 2006). Unlike these approaches, the CSI assumes that the analyst has no known groupings (similar to the "actuals" or "y-values" in regression models) with which he/she can compare peer group results. The CSI cannot validate its results by comparing a trial result with known classifications (even for a sample of entities). In other words, there is no pre-existing benchmark data.

Within traditional cluster analysis, the analyst has tools to diagnose the quality of his/her classifications or peer groups. For example, he/she could examine the proximity measures between the different members (i.e., review the distance matrix) or the level of splitting that occurs in a hierarchical cluster analysis which the dendrogram and agglomeration schedule facilitate (Hair, et al., 2006; Everitt, et al, 2001). Other tools include the comparison of the statistics (i.e., means and standard deviations) for each peer group to ascertain homogeneity within each peer group and heterogeneity between the peer groups (Klastorin, 1983). The methods that an analyst may apply are numerous, and these methods can become fairly complex (Xu & Wunsch, 2009; Fielding, 2007; Tibshirani, Guenther & Hastie, 2001; Punj & Stewart, 1983; Rand, 1971). However, traditional approaches avoid the question of how much effect method bias has upon a specific member or entity of a classified population.

Assumptions

In using the CSI, analysts must accept a set of major assumptions. These assumptions are important because any neglect or rejection of them will mean that the CSI is not an appropriate procedure for a given situation. Because cluster analysis functions to classify members of a population into mutually exclusive groups, we will refer to cluster analyses below as *classification efforts* to help emphasize the generalizability of the CSI to other methods (such as classification and regression trees and multiple discriminant analysis).

1. A given classification effort can produce a different set of peers for an institution if the analyst executes different classification methods on the same data set (i.e., a fixed set of cases and variables).
2. The variables in the classification scheme are relevant to the classification purpose.
3. The analyst does not have pre-existing benchmark data by which to quantify accuracy of classification.
4. Each classification method or alternative has equal validity for the determination of an institution's peers. That is, we accept that each method is just as meaningful or credible as the next one for a given classification effort.
5. The primary interest of the analyst is the level of variation in the identification of a given institution's peers and not the validation of one method's results in comparison to any alternative method.
6. An analyst can use the results of one classification method as a "baseline" or reference model with which to calculate "difference" implied by any alternative methods.
7. The population from which we wish to find peers is fixed, known, and relatively small.
8. Each classification method must classify each case in the population (that is, cases cannot be excluded from a classification method).
9. Data users can accept that a peer group can be a "fuzzy" set, so to speak, rather than a "black-and-white" specification of peer institutions (i.e., distinctly separate groups with no overlap).

Calculating the CSI

The overall concept of the CSI is simple. Different classification or clustering methods will usually produce different peer group members for a specific institution, for example "Best U." The number of different peers that the alternative methods identify for Best U represents the effect of method bias in the classification of Best U, given the variables considered in the data analysis. That is, the method bias is measured by summing the number of "new" peer group partners for Best U in comparison to its smallest peer group result from the different methods. In the extreme case, Best U may be an outlier—a peer group size of one. There may be a special situation where different classification methods produce two minimum peer group sizes containing different sets of institutions, but there is still a tie in terms of numbers. Because the handling of a tie is a technical matter, this special situation is addressed in the Appendix.

If it is found that the classifications for an institution have high variability in their memberships, then it is inferred that the institution's classification is highly sensitive to the method chosen to classify it. Conversely, if low variability is found, the institution's classification has low sensitivity to choice of method. In other words, if the peers identified with different classification methods vary substantially for a given institution, then there is evidence that a given institution's peer group is highly sensitive to method bias.

An institution's classification that has a low level of method bias would merit a higher level of decision-making confidence than one that has a high level of method bias. Finally, to make the CSI into an index so that it will take any value between and including zero and one, we divide the number of additional institutions assigned from alternate methods by the total number of institutions that were possible "partners" of a specific institution. With this index, a value of one represents maximum variability in the classification of an institution, and a value of zero represents minimum variability. Figure 1 summarizes this concept.

In Figure 2, algebraically we have the following basic formula for the CSI:

In Figure 2, the phrase "count of institutions found as Best U peers across all methods" refers to the number of different institutions that the grouping methods may identify as peers for Best U. If two grouping methods identify a hypothetical college, such as Central U as a Best U peer, the equation will only count Central U once for the numerator in Figure 2. The "minimum group count among all methods" will always be equal to or greater than one because a college without any identified peers will constitute a group of one institution (by itself) as a matter of definition. So that this equation will always result in a non-negative number, 1 is added to the "count of institutions found as Best U peers across all methods." Therefore, a college that has no peer institutions identified for it in an analysis that uses several different clustering methods will receive a CSI of zero (because the numerator will be zero divided by a positive number).

The calculations with a simulated data set are in the following section. In the hypothetical situation depicted in Figure 3, there are 10 institutions (A – J) that an analyst classified via 4 different methods (W-Z), resulting in 3 different peer groups (cell values green, red, or blue) per method.

Thus, for Institution A, Method W identified it as an outlier because it has no peers (only 1 row in the column for Method W has a "green" in it). Method X differs in that it identifies Institution B as a peer for Institution A (denoted by the "green" in the cell for the intersection of the column for Method X and the row for Institution B). Method Y agrees with Method X in terms of the peer for Institution A (but note that Method Y disagrees with Method X in terms of the peers for Institutions E-J). This occurs primarily because Method Y only produces 2 peer groups from the 10 institutions, whereas the other 3 methods produce 3 peer groups. This is an important point because different classification methods can produce not only different peer members for a given institution, but they can also produce different numbers of groups or clusters from the same set of institutions. The CSI preserves its interpretation regardless of changes in the number of peer groups that each classification method may define.

Figure 1. The CSI Range of Values

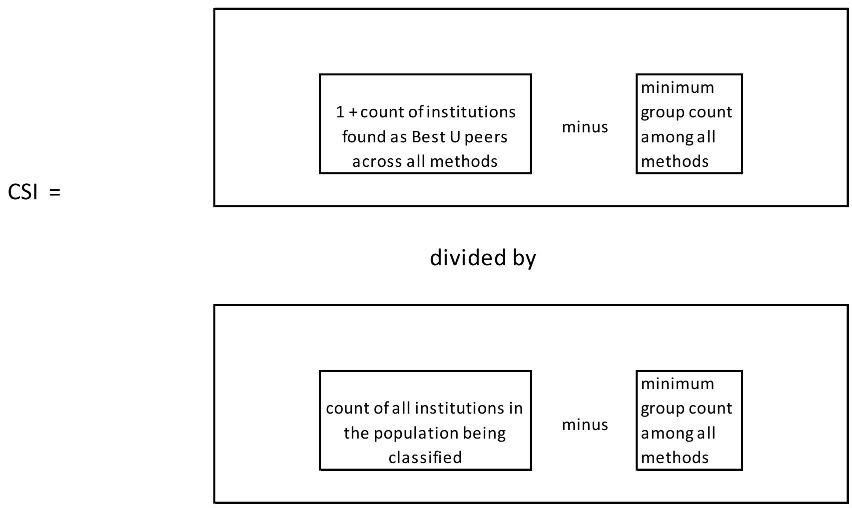

Figure 2. Calculating the CSI

The peer group results are isolated for Institution A by creating a table (see Figure 4) that displays only the peers that the four methods have identified for Institution A.

To calculate the CSI for Institution A, begin by finding the smallest grouping containing A, given the simplified table in Figure 4. This is the cluster of one institution produced by Method W. In Method W, Institution A is an outlier (by itself). Next count the number of **new** institutions that the other methods identify as peers for Institution A. Method X identifies a peer (Institution B), and then add 1 to the count. In terms of classifying Institution A, Method Y duplicates Method X (renaming Institution B as A's peer), so the count stays at 1. Method Z differs from the other three methods by adding Institution J as a peer for Institution A, so the count is increased by 1, ending with a sum of 2.

This sum of 2 does not account for the number of institutions in this data set that could have been identified as peers for Institution A. So this sum is not comparable to the corresponding sums that are calculated for any other institution in this population. To make this sum comparable for different institutions in this classification exercise, divide this sum by the number of new institutions that could potentially have been identified as peers for Institution A. Since the smallest peer group result for Institution A had a count of 1 (A by itself), the denominator in the division is 9. Evaluating the expression 2/9 results in a CSI of 0.222 for Institution A in this exercise. This calculation is summarized in Figure 5 below.

Institution	Method W	Method X	Method Y	Method Z
A	green	green	green	green
B	red	green	green	green
C	red	red	red	red
D	red	red	red	red
E	blue	blue	red	blue
F	blue	blue	red	blue
G	blue	blue	red	blue
H	blue	blue	red	blue
I	blue	blue	red	blue
J	blue	blue	red	green

Note: Cell values identify the different clusters (or peer groups) that result from different clustering methods.

The four colors represent peer groups of the 10 institutions listed. Each column represents a different methodology and how each institution was classified.

Figure 3. Hypothetical Results as Four Different Classification Methods on a Popluation of 10 Institutions

Institution	Method W	Method X	Method Y	Method Z
A	green	green	green	green
B		green	green	green
C				
D				
J				green

Figure 4. InstitutionsIdentified as Peers for Institution A

Some characteristics of the CSI deserve mention. As the calculations for Institution A demonstrate, the CSI calculation can be tedious, especially if the number of peer members and the population undergoing classification are large. Within one data scenario that uses multiple classification methods, different institutions can have widely divergent CSI's. Even if an analyst assumes a different data scenario, some CSI's may stay the same; it all depends upon the specific data set used. The CSI can have a value of zero—signifying no classification variability or method bias in the identification of an institution's peers. The CSI can obtain a value of 1.000 to indicate maximum variability in classification. It would be unusual to obtain a CSI of 1.000 with a real data set because this would indicate that an institution somehow resembles all the other entities in a cluster analysis while somehow causing different clustering methods to totally disagree in their results.

An Example with Real Data

A brief examination of how the CSI operates with a real data set may help readers appreciate the properties of the CSI. This exercise includes 109 community colleges in the state system in 2007, using two key institutional variables (enrollment size during the fall term of 2005 and the economic backgrounds of the historical student enrollees). Because of missing data, this example will only classify 108 institutions. To simplify the example and to demonstrate a potential real-world analytical environment, this exercise will only use 3 different classification methods to produce 3 sets of peer group results. In this example squared Euclidean distance with average-linkage is used as the option for one classification method. Squared Euclidean distance with Ward's method is the option for the second classification method. The third method used is the Minkowski distance with Ward's method as the option and is labeled as "Ward's Method II" in the figure below. Figure 6 below displays an excerpt of the results as they pertain to the peers identified for Palomar College (chosen for convenience). Unlike our prior tables, peer membership is denoted with Palomar with an "X" in the appropriate cell of this table.

These cluster results appear to agree well in terms of the peers identified for Palomar. To calculate the CSI for Palomar, proceed in the following manner. Ward's Method II has the smallest peer group (n = 11) for Palomar among the methods used here. Ward's Method (the middle column) adds Long Beach, East Los Angeles, El Camino, and Sacramento City to the minimum peer group defined by Ward's Method II. So the subtotal stands at 4. The Average Linkage Method adds Moorpark, making the total of added peers equal to 5. Since there were only 11 colleges in the smallest defined peer group, we find that 97 colleges (or 108 minus 11) could have been potentially identified by the other two clustering methods in this example. So Palomar's CSI is 5/97 or 0.052. In this situation, Palomar's classification has very little variability, and an analyst should have high confidence in the identification of Palomar's peers on these institutional variables. Its peer group classification is robust. Of course, whether or not these same peer colleges are peers of Palomar on other institutional variables remains unanswered here.

Using the CSI

The author has shown how to calculate the CSI for a set of cluster results, and now will explain how the analyst can use these indices. Of course, most analysts will want to recognize any benefit of the CSI before they even consider going to the effort to calculate this index. The following discussion will attempt to identify ways to improve analyses and decision-making from classification efforts. Practically speaking, analysts should encounter low CSI numbers if the two following conditions exist:

1. True sub-groups exist in a given population, and
2. The cluster analysis (or other classification method) has used a set of variables that effectively capture the true differences between an institution's peers versus its non-peers.

A set of high CSI numbers for a classification effort would indicate that there is a problem with either condition 1 or condition 2, or with both conditions. Under any of these interpretations of the CSI, the analyst can use the index as a diagnostic tool in his/her standard protocol for finding peer groups.

A CSI of 0.50 indicates that the difference between an institution's smallest peer group (derived from one clustering method) and its largest peer group covers half of the population that could have been classified as peers after considering the smallest peer group. In most situations, an analyst would consider such a CSI level as evidence of a rather tenuous identification of an institution's peers. Likewise, a CSI of 0.333 tells us that 1/3 of the remaining institutions (beyond those identified as the smallest peer

$$\text{CSI} = \left\{ \begin{array}{c} (1+2 \text{ new institutions} = 3) \text{ minus (minimum group count} = 1) \\ \div \\ (10 \text{ total institutions} - 1 \text{ minimum group count}) = 9 \end{array} \right\} 2/9 = 0.222$$

Figure 5. Caculation of the CSI for Institution A

Institution	Average Linkage Method	Ward's Method	Ward's Method II
Palomar	X	X	X
American River	X	X	X
Sacramento City	X	X	
Santa Rosa	X	X	X
Diablo Valley	X	X	X
San Francisco	X	X	X
De Anza	X	X	X
Moorpark	X		
El Camino	X	X	
East L.A.	X	X	
Pasadena	X	X	X
Santa Monica	X	X	X
Long Beach	X	X	
Mt. San Antonio	X	X	X
Saddleback	X	X	X
Riverside	X	X	X

Note: "X" denotes membership in the peer group for Palomar College.

Figure 6. Results from Three Clustering Methods with Actual College Data

group for that institution) were also possible peer institutions according to alternate clustering methods. This seems more acceptable than a CSI of 0.50 for decision-making, but 0.333 may still seem too tenuous for a policy decision that can profoundly affect people. From a commonsense perspective, 0.10 would seem to be an acceptable level of variability (or robustness) in peer member identification. But given the widely varying costs of error for each peer group analysis, it is obvious that we do not have any fixed thresholds for the CSI to simplify a judgment that a particular peer grouping is usable or not. Consequently, until more analysis occurs on the CSI, analysts will need to rely on personal judgment to decide the acceptability of a particular CSI level.

Note that an institution's CSI for a given peer grouping analysis will not equal that same institution's CSI for a second peer grouping analysis if there are a different number of institutions in the second analysis. As demonstrated with the case of Palomar, the CSI for one institution will vary with the counts of different populations that an analyst may use in a set of peer grouping scenarios. Figure 7 below shows the results of scenarios that involve the same pattern of group change for Palomar, but with different population counts in the peer grouping effort. As the population count rises, the CSI falls, despite the use of the same peer group membership for Palomar in each scenario. This means that analysts should avoid comparing the CSI's from peer grouping results that involve significant differences in the size of the populations. Practically speaking, however, this sensitivity to changes in population size should not diminish the utility of the CSI as a diagnostic tool. Analysts generally work with a finite, or fixed, population in their peer grouping scenarios,

Population Count in a Scenario	CSI
25	0.36
50	0.13
100	0.09
108	0.05
200	0.03
300	0.02

Figure 7. The CSI in Scenarios with Different Population Counts

and the differences in population will not alter the CSI much when a difference in the population counts is small (a scenario that may occur when one new institution joins a system or when one exits).

Conclusion

The CSI is a diagnostic tool that should receive further testing and development. This index has not been used with a wide spectrum of real-life data situations. It is quite possible that application of the CSI may lead to improvements. The CSI may have practical benefits for analysts and decision-makers because it has intuitive appeal, relatively simple arithmetic in computation, and generalizability to any kind of classification effort. It applies to tools including, but not limited to, cluster analysis. The intuitive appeal of the CSI comes largely from the CSI's focus upon the range of new peer institutions that the analyst can identify with alternative methods of clustering or classification. The simplicity of computation may become tedious with more than a few classification methods in a situation or with more than a couple dozen institutions in the population to be classified. However, some programming via a database or a spreadsheet program could reduce the computational burden.

A major advantage of the CSI is its generalizability, which has two major dimensions to it. Because the CSI only operates upon the results of a classification effort, rather than upon the processes of classification, it can be used with classification processes that involve people (opinion surveys) or statistical/computational tools (including the suite of data mining tools). Within the realm of cluster analysis, the CSI's generalizability could easily be broadened if the proximity measures and the clustering algorithms for the set of test peer groups were to be randomly selected. Secondly, the CSI can be applied to any kind of population; the entities in the classification could be cities, individual personalities, consumer products, or concepts. The focus upon the peer grouping of postsecondary institutions in this paper offers only a brief glimpse of the CSI's utility for a universe of other kinds of classification situations.

References

Aldenderfer, M.S., & Blashfield, R.K. (1984). *Cluster analysis*. Thousand Oaks, California: Sage.

California Community Colleges, Chancellor's Office. (2009, February). Focus on results. Retrieved on 2/02/2009_from http://www.cccco.edu/Portals/4/TRIS/research/ARCC/arcc_2008_final.pdf

Carnegie Foundation for the Advancement of Teaching. (2007, October). The Carnegie classification of institutions of higher education. Retrieved on February 2, 2009 from http://www.carnegiefoundation.org/classifications/

Everitt, B.S., Landau, S., & Leese,M. (2001). *Cluster analysis*. London: Arnold.

Fielding, A.H. (2007. *Cluster and classification techniques for the biosciences*. New York: Cambridge.

Frank, R.E., & Green, P.E. (1968). Numerical taxonomy in marketing analysis: a review article. *Journal of Marketing Research*. 5, February: 83-98.

Funkhouser, G.R. (1983). A note on the reliability of certain clustering algorithms. *Journal of Marketing Research*. 20, February: 99-102.

Goan, S.K., & Cunningham, A.F. (2007). Differential characteristics of 2-year postsecondary institutions (NCES 2007-164). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Green, P.E., & Rao, V.R. (1969). A note on proximity measures and cluster analysis. *Journal of Marketing Research*. 6, August: 359-364.

Han, J., and Kamber, M. (2001). *Data mining:concepts and techniques*. San Francisco: Morgan Kaufmann.

Hand, D., Mannila, H., and Smyth, P. (2001). *Principles of data mining*. Cambridge, Massachussets: MIT.

Lorr, M. (1987). *Cluster analysis for social scientists*. San Francisco: Jossey-Bass.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., & Tatham. (2006). *Multivariate data analysis*. Upper Saddle River, New Jersey: Prentice-Hall.

Hom, W.C. (2008). Peer grouping: the refinement of performance indicators. *Journal of Applied Research in the Community College*. 16(1), 45-51.

Hurley, R.G. (2002). Identification and assessment of community college peer institution selection systems. *Community College Review*. 29, 1-27.

Jain, A.K. & Dubes, R.C. (1988). *Algorithms for clustering data*. Englewood Cliffs, New Jersey: Prentice-Hall.

Ketchen, D.J. & Shook, C.L. (1996). The application of cluster analysis in strategic management research: an analysis and critique. *Strategic Management Journal*. 17, 441-458.

Kettenring, J.R. (2006). The practice of cluster analysis. *Journal of Classification*. 23, 3-30.

Klastorin, T.D. (1983). Assessing cluster analysis results. *Journal of Marketing Research*. 20, February: 92-98.

McCormick, A.C., and Cox, R. D. (2003). Classifying two-year colleges: purposes, possibilities, and pitfalls. In A.C. McCormick & R.D. Cox (Eds.) *New directions for community colleges*. (pp.7-16). San Francisco: Jossey-Bass.

McCormick, A.C., & Zhao, C. (2005). Rethinking and reframing the Carnegie Classification. *Change*. September/October, 51-57.

Phipps, R.A., Shedd, J.M., and Merisotis, J.P. (2001). *Classification system for 2-year postsecondary institutions, a*. (NCES 2001-167). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Punj, G., & Stewart, D.W. (1983). Cluster analysis in marketing research: review and suggestions for application. *Journal of Marketing Research*. 20, May: 134-148.

Rand, W.M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*. 66 (336), 846-850.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society*, Series B. 63,2, 411-423.

U.S. Department of Education, National Center for Education Statistics. (2007, December). *IPEDS peer tool and peer analysis system*. Retrieved on February 2, 2009 from http://nces.ed.gov/ipedspas/Xu, Rui, and Wunsch, D.C. (2009). *Clustering*. Hoboken, New Jersey: Wiley & Sons

---

Willard Hom *is the Director of Research & Planning for Chancellor's Office, California Community Colleges in Sacramento, California. He has served as a researcher for the California Department of Health, the California Employment Development Department, and the Chancellor's Office during the past three decades. He also serves as the President of the National Community College Council for Research & Planning and the President of the California Association of Institutional Research. His research interests include performance measurement, surveys, and consumer behavior.*

Acknowledgments

Catharine Liddicoat, Alice van Ommeren, and LeAnn Fong-Batkin provided helpful comments for the initial drafts of this paper.

Appendix

Handling Ties in the Peer Groups for the Minimum Count of Members

There may be a special situation where different classification methods produce two minimum peer group sizes containing **different** sets of institutions, but we still see a tie in terms of numbers. To explain how to handle this special case for computation of the CSI, we use the data below in Figure A-1. The numeral "1" in a cell indicates that the institution shown in that row is identified by the method shown in that column as a peer member for Institution A. Here, both Methods W and X result in peer groups of size two. However, Method W has Institution A with Institution B while Method X has Institution A with Institution E.

In this situation, we simply choose one of the methods as a "base" and count the additional peers identified in the other minimum-sized peer group as part of the variation from the smallest peer group count that we derive. To walk through our tie example then, we arbitrarily make Method W our "base" minimum peer group. Method X, the other method with a minimum peer group of two, added Institution E as a peer to Institution A. Methods Y and Z added only one more institution as a peer (i.e., Institution C).

Thus we have a variation of two from our chosen minimum (or base) of two in Method W. We have a CSI of 2/8, or 0.25, because there were only eight potential institutions (after excluding Institutions A and B) that could have become peers of A. Note that if we had chosen Method X (instead of Method W) as the base, the result would also be 0.25. So, the choice of either Method W or Method X as a base makes no difference to the CSI in this special kind of a tie in the minimum peer group.

Institution	Method W	Method X	Method Y	Method Z
A	1	1	1	1
B	1		1	1
C			1	1
D				
E		1		
F				
G				
H				
I				
J				
Column Count	2	2	3	3

Figure A-1. CSI for Institution A When a Tie Exists for the Smallest Peer Group