PEER GROUPING—SIMPLIFIED

A Note from Accountability Reporting for the Community Colleges (ARCC)

When analysts (i.e., institutional researchers) search for best practices or try to compare the performances of various colleges, they often try to compare each college to its peers. This prevents an analyst from comparing "apples to oranges." If analysts want to compare the performance of a community college (say a college named Alturas) on a function such as student persistence, then they should find those colleges that match Alturas on factors that affect student persistence. If the analysts find that environmental factors, such as average age of the enrollment and average unit load (an indicator of the level of part-time enrollment), affect student persistence, then the identification of peers for Alturas implies a search for colleges that have levels of student age and student unit load that resemble those at Alturas.

In this hypothetical scenario, analysts need to collect, from each college in a region, data on the age of the enrolled students and the number of units that each student takes. By comparing each college's numbers on these two factors, analysts can identify those colleges that seem to have very similar levels of the factors. If the analysts only need to compare a few colleges and if they only need to compare them upon two environmental factors, then the analysts could identify colleges similar to Alturas (i.e., the peers for Alturas) by simply examining a two-way graph of each college's numbers (with average student age on the vertical axis and average unit load on the horizontal axis). The points on the graph represent colleges. Those points that appear to "bunch" closely around Alturas would obviously qualify as its peers. On the other hand, points that seem to have more "distance" from Alturas on this graph, in comparison to those closely bunched around Alturas, would not be peers for Alturas.

If the analysts need to identify the peers for another hypothetical college, Upper Gulch CC, then they would repeat this search for the points that bunch around the point that represents Upper Gulch CC on this two-way graph. Naturally, if the analysts had to identify the peer groups for a population of twenty colleges, they would need to repeat the process that they used for Alturas and Upper Gulch CC—eighteen more times.

In the ARCC project, analysts found that student persistence involved three environmental factors, and the ARCC needed to identify peers for 109 colleges. With three factors, it becomes nearly impossible to identify "bunchings" of points on a three-dimensional (3-D) graph, and to do this 109 times would involve a good deal of time. Furthermore, the analyst who must visually inspect a graph for bunchings is likely to have some bias about which points appear closest compared to other points. Even measuring with a ruler will create some error.

To remedy these drawbacks to a visual inspection of graphs, the ARCC researchers used a mathematical tool known as *cluster analysis*. Cluster analysis accomplishes the same goal of "inspecting" a graph of points for "bunchings" or clusters. However, cluster analysis does the inspection much more quickly than a human can do this and with considerably more precision. Cluster analysis uses the computed "distances" between the points on the graph rather than a physical measurement of space between the points on a graph printed on paper. In a way, cluster analysis is like automated surveying of points on a piece of land. This mathematical tool uses a computer program to "measure distances" between points in a "virtual" space, and it tells you which points in this virtual space are really close to one another. After the program identifies a college's peers (those points that lie closest to a college), the program then defines those points that lie too far away to be part of that college's peer group. It's that simple.