WHITE PAPER SERIES

Selecting Peer Institutions Using Cluster Analysis - Fall, 2018

Decision Support and Institutional Research

Austin Peay State University

About the Author

Dr. Andrew L. Luna, is Executive Director of Decision Support and Institutional Research. He has served over 30 years in higher education, with 21 of those years in institutional research. He has published research studies on many topics including salary studies, assessment, market research, and quality improvement. Dr. Luna received his Ph.D. and M.A. degrees in higher education administration and his M.A. and B.A. degrees in journalism, all from the University of Alabama.

Table of Contents

Executive Summary.....1

Introduction.....2

IPEDS Initial Institutional Screening.....5

Running the Cluster Analysis Procedure.....7

Determining Fit and Reliability of Model.....8

Results.....10

Appendix I.....14

Appendix II.....15

Appendix III.....17

Appendix IV.....19

EXECUTIVE SUMMARY

Sensing a need to update Austin Peay State University's peer institution list, the Office of the Provost and Vice President of Academic Affairs charged Decision Support and Institutional Research (DSIR) with the task of creating a more scientific and reliable method for selecting APSU's peers.

The method used is referred to as cluster analysis, which is defined as an exploratory data analysis technique for classifying and organizing data into meaningful clusters, groups, or taxonomies by maximizing the similarity between observations within each cluster. The purpose of cluster analysis is to discover a system of organizing observations into groups where members of the groups share properties in common.

The first iteration of the project was not satisfactory because the base set of institutions provided by a federal government application was incomplete. This second iteration established a base using all master-level (large programs) institutions with enrollments between 6,000 and 16,000 that were not Historically Black Colleges and Universities, land grant, or tribal institutions (n=73). Based upon the analysis, DSIR recommends including institutions from two closely-related clusters and who share in the same regional accreditation in its peer list:

1. Columbus State University (Georgia)
2. Jacksonville State University (Alabama)
3. McNeese University (Louisiana)
4. Morehead State University (Kentucky)
5. Murray State University (Kentucky)
6. Radford University (Virginia)
7. Texas A&M International
8. The University of Tennessee - Chattanooga
9. The University of Tennessee - Martin
10. The University of Texas at Tyler
11. University of Houston - Clear Lake (Texas)
12. University of Houston - Victoria (Texas)
13. University of North Alabama
14. University of North Carolina at Pembroke

Further analysis could use different comparator variables, changing the initial screening to include institutions within a specific region, or choosing institutions from higher-ranked clusters for possible inclusion in an aspirational peer list.

INTRODUCTION

Within the current state of higher education, colleges and universities must strive to be competitive in both the quality of education they offer as well as the cost of attendance. At the same time, higher education is being held more accountable by federal and state governments, as well as by the communities they serve. This accountability varies broadly by legislative bodies, governors' offices, faculty committees, federal mandates, students and other constituencies. Therefore, the use of comparator institutions as a reference point within higher education has become common practice.

The use of peer comparator institutions allows administrators to compare both the quality and quantity of academic programs and delivery methods, as well as institutional expenditures and revenues. Comparisons like these allow for more focused strategic and long-range planning strategies in order to meet goals and objectives.

When identifying peers, it is important to understand the focus for the comparison group, as more than one set of peer groups may be utilized by an institution. There are various kinds of peers, such as:

- **Comparable:** Similar institutional level (two-year vs. four-year), control (e.g. private not-for-profit vs. public) and enrollment profile characteristics.
- **Aspirational:** Institutions with similar institutional characteristics yet are significantly different in several key performance indicators, such as significantly higher graduation rates or endowments.
- **Competitors:** Based on cross applications, institutions may have different institutional characteristics, yet a significant percentage of the institution's applicants choose to attend another institution.
- **Consortium:** Institutions belonging to a consortium for a common purpose and/or to share data may be another peer group for review.

The purpose of this study is to identify those institutions that are comparable to APSU. These peer institutions share the same basic Carnegie Classification (e.g. Master's Institution vs. Associate of Arts), in addition to similar graduation rates and enrollment mix (e.g. percent full-time vs. part-time).

Since 2013, APSU created its list of peer institutions for use as comparators in the Integrated Post Secondary Educational Data System (IPEDS) Institutional Feedback report. This peer list (below) was created from institutions located only in Tennessee:

1. East Tennessee State University (Johnson City, TN)
2. Middle Tennessee State University (Murfreesboro, TN)
3. Tennessee State University (Nashville, TN)
4. Tennessee Technological University (Cookeville, TN)
5. The University of Tennessee-Chattanooga (Chattanooga, TN)
6. The University of Tennessee-Martin (Martin, TN)
7. University of Memphis (Memphis, TN)

APSU also worked with the National Survey of Student Engagement (NSSE) to develop an extensive set of peer institutions based on type, size, location, and Carnegie Classification. This list was somewhat broad and over-inclusive in order to allow APSU to develop a group of adequate size to compare results of the NSSE survey.

Additionally, APSU partnered with the College and University Professional Association (CUPA) to conduct various salary studies. As part of this process, CUPA developed a set of peer institutions as a base for salary comparisons. While many institutions within the CUPA list are comparable to APSU, the list of 68 institutions (**Appendix III**) is also broad in that it includes research institutions, institutions with over 15,000 enrollment, as well as Historically Black Colleges and Universities (HBCU).

Sensing a need to create a peer list that is more similar to APSU, the Provost and Vice President for Academic Affairs' Office charged DSIR with the task of creating a more scientific and reliable method for selecting APSU's peers.

The process of utilizing statistical methodologies in the identification of peer institutions began more than 20 years ago (Terenzini, et al., 1980; Teeter & Brinkman, 1987; and McLaughlin & McLaughlin, 2007). The overall goal during this time has been to identify appropriate methods for comparing the performance of a reference institution relative to a group of similar institutions, and to make goal and outcome decisions concerning the reference institution based on the performance of the comparator institutions.

"The process of utilizing statistical methodologies in the identification of peer institutions began more than 20 years ago."

While the use of statistical methodologies supports scientific objectivity, their complexity often makes them difficult to understand by the end user. Other studies have also indicated that these types of methodologies inherently contain statistical error due to the additive and multiplicative attributes of the procedures used (McLaughlin & McLaughlin, 2007). It is, therefore, recommended that the institution not rely solely on the outcome of a statistical peer analysis. Rather, the data from the analysis should be used in conjunction with other knowledge gained.

This study used cluster analysis, which is defined as an exploratory data analysis technique for classifying and organizing data into meaningful clusters, groups, or taxonomies by maximizing the similarity between observations within each cluster. The purpose of cluster analysis is to discover a system of organizing observations into groups where members of the groups share properties in common. The goal of this analysis, therefore, is to sort variables into groups or clusters so that the degree of association or relationship is strong between members of the same cluster and weaker between members of different clusters.

"...cluster analysis, [is] defined as an exploratory data analysis technique for classifying and organizing data into meaningful cluster, groups, or taxonomies..."

The appropriate cluster algorithm and parameter settings depend on the individual data set and intended use of the results. Furthermore, cluster analysis is an iterative process of knowledge discovery and optimization to modify data processing and model parameters until the result achieves both the preferred as well as appropriate properties.

The choice of methods used for cluster analysis depends on the size of the data set as well as the types of variables used. In this study, hierarchical clustering is more appropriate because the data set is small. The steps in obtaining and preparing the data for cluster analysis are as follows:

- Screen institutions to determine what type and size of institution will be used in the analysis
- Choose variables to download from IPEDS that will be used in the analysis
- Standardize all quantifiable variables that will be used in the analysis
- Run the cluster analysis procedure
- Determine the fit and reliability of the model
- Identify those institutions that are within the same cluster as APSU

IPEDS INITIAL INSTITUTIONAL SCREENING

To start the process of determining institutional peers, an initial reference group was established. Larger research institutions, two-year colleges, and specialty institutions with a significantly different role, scope, and mission than APSU were screened out. This screening process was first generated through the grouping procedure found within the IPEDS Data Center. However, this procedure only yielded 41 institutions and a few of these included institutions that APSU should omit (HBCUs and research institutions). Therefore, a new list of institutions was generated through the IPEDS system by choosing only public 4-year institutions with a Carnegie Classification of Masters - Larger programs, and total enrollment between 6,000 and 16,000.

"Larger research institutions, two-year colleges, and specialty institutions with a significantly different role, scope, and mission were screened out."

From these criteria, a total of 73 institutions were included for the cluster analysis. A listing of all institutions used in this study can be found in **Appendix IV**. From these institutions, specific variables were chosen to be used in the cluster analysis procedure.

Choosing Variables to Use in the Analysis

Once the initial 73 institutions were selected, a total of 12 selected variables were downloaded from the IPEDS Data Center for each institution. These variables were selected by DSIR following an extensive literature review process on what key variables are factors in determining institutional role, scope, and mission. The variables selected are listed below:

1. Undergraduate enrollment for latest fall semester
2. Graduate enrollment for latest fall semester
3. FTE for latest academic year
4. Six-year graduation rate based on the IPEDS defined freshman cohort
5. Total operational revenues
6. Tuition and fees as a percent of operational revenues
7. State appropriations as a percent of operational revenues
8. Total expenditures
9. Instructional costs as a percent of expenditures
10. Endowment assets per FTE
11. In-state tuition and fees on-campus
12. Out-of-state tuition and fees on-campus

Once the variables for the study were pulled, it was determined that data values of the 12 variables were present in each of the 73 institutions. Therefore, none of the original 73 institutions were removed based upon insufficient data.

Given that the raw data pulled from IPEDS for this analysis significantly vary, all of the variables were then standardized for use in the analysis. It should be noted that using variables without standardizing them can give those variables with larger values and ranges greater importance in the analysis. Standardizing the variables remedies this issue. The standardization used in this study is reviewed in **Appendix I** of this report.

"It should be noted that using variables without standardizing them can give those variables with larger values and ranges greater importance in the analysis."

RUNNING THE CLUSTER ANALYSIS PROCEDURE

The objective of cluster analysis is to group observations of interest into clusters so that those observations within each group are similar inside the group while each group stands apart from each other.

Take, for example, a group of people who are inside a stadium. As one large group, there exists a lot of variability and difference. If the larger group was parsed out by certain key variables or attributes, those people who were married, earned high income, and had multiple children would fall into one cluster or group while individuals who had lower incomes, were single, and had no children would be placed into another group. In this example, individuals having similar attributes would be in the same cluster while those who were different would fall into another cluster.

"While there are numerous ways in which clusters may be formed, hierarchical clustering is one of the most straightforward methods."

While there are many ways to run a cluster analysis, there are two basic fundamental methods of hierarchical (systematic) analysis. The first method involves forming as many groups as there are observations and the systematically merging observations in order to reduce the number of groups. This method is call agglomerative. The second basic method is called divisive in that it groups all observations into one cluster and then separates the observations into like groups. This study uses the former, agglomerative, method.

It is important, however, to note that whatever method is used, the risk of under or over specifying the model may occur. For example, if there are 80 observations, there is clearly little benefit in grouping all 80 into one group or, likewise, to place each observation into one of 80 clusters.

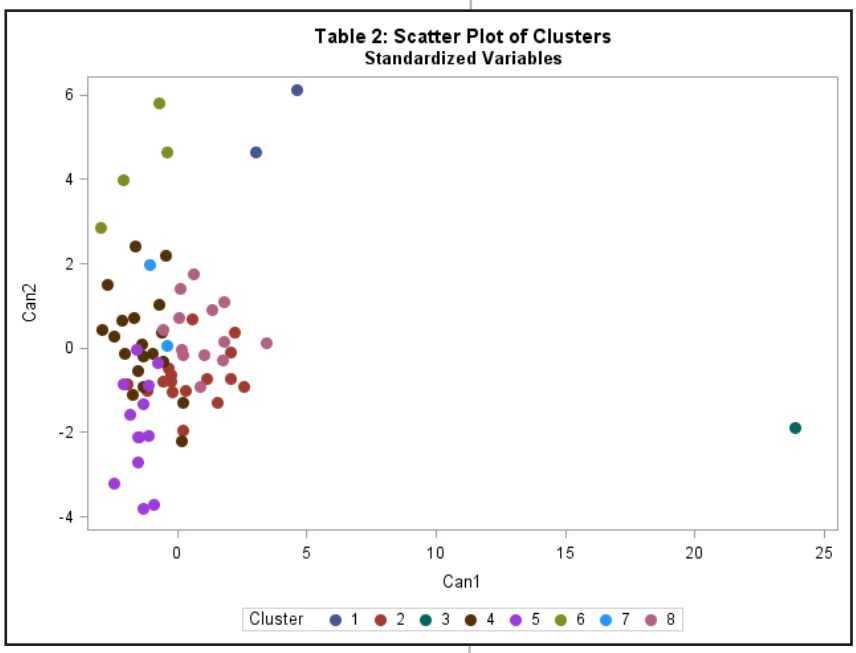
The mechanics of running a cluster analysis involves determining distance of each variable or attribute within an observation and grouping similar distances together. For a more detailed explanation of cluster analysis, please refer to **Appendix II** of this report.

DETERMINING FIT AND RELIABILITY OF MODEL

After the cluster analysis procedure determined that the 73 institutions could be reduced into eight unique clusters, a canonical discriminant analysis was run to create grouped variables for use in a scatter plot in order to determine where each of the clusters fall. Canonical discriminant analysis is used to find a linear combination of features which characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier or, more commonly, for dimensionality reduction before later classification.

In essence, the canonical discriminant analysis determines distances between one or more quantitative variables and then determines the relationship between the quantitative variables and a set of classification variables to place observations into clusters so that every observation belongs to one and only one cluster.

The first canonical correlation is the maximum correlation that can be obtained between a linear combination of one set of variables and a linear combination of another set of variables. The second canonical correlation is the maximum correlation that can be obtained between linear combinations of the two sets of variables subject to the constraint that these second linear combinations are orthogonal (independent/uncorrelated) to the first linear combinations. The second canonical variable provides the greatest difference between group means while being uncorrelated with the first canonical variable.



Within this study, the first two canonical correlations explain about 96% of the variation in the study, so plotting the first canonical correlation against the second should give a good indication where the clusters fall and how closely related they are to each other. Following the FASTCLUS procedure in SAS, the first canonical variable was

plotted against the second canonical variable. Together, these variables indicate where the various clusters reside, how widely distributed they are, and how close they are to each other. As can be seen in **Table 2**, all of the clusters are distinct, albeit they are close together with clusters 2 (red), 4 (brown), and 5 (purple) overlapping in some places. Clusters 1 and 3 contain only one or two institutions and, therefore, are considered outliers. APSU is within cluster 5.

In determining which cluster the home institution resides, it is also important to note a couple of other pieces of information from the scatter plot. First, some clusters are very close to the home institution's cluster and, in some cases, may actually be intertwined within the home cluster. If there was a desire to remove institutions from the home cluster due to geographic reasons, other institutions could be used from clusters that are close to the home cluster.

It should be noted that in determining the final peer group, some of the institutions within the home cluster that reside outside of the southeast may be replaced with southeastern institutions residing in clusters close to or intertwined within the home cluster.

Another piece of information to consider while observing the scatter plot is to see the relationships of the other clusters to the home cluster. For example, in observing clusters ranked above the home cluster, the institution may want to look at institutions within these clusters as possible aspirational peers.

"It should be noted that in determining the final peer group, some of the institutions within the home cluster that reside outside of the southeast may be replaced with southeastern institutions residing in clusters close to or intertwined within the home cluster."

RESULTS

The FASTCLUS procedure within SAS indicated that the 73 institutions would best be divided into eight clusters with APSU residing in cluster 5. The results also indicate that cluster 4 was close to the APSU cluster. It should be noted that clusters 1 and 3 contain only one or two institutions each and should be considered outliers. Those institutions within cluster 5 were:

1. Columbus State University (Georgia)
2. Emporia State University (Kansas)
3. Fitchburg State University (Massachusetts)
4. Indiana University-Southeast
5. Jacksonville State University (Alabama)
6. Northeastern State University (Oklahoma)
7. SUNY College at Plattsburg (New York)
8. Southern Utah University
9. The University of Tennessee - Martin
10. The University of Texas at Tyler
11. University of Central Missouri
12. University of Houston - Clear Lake (Texas)
13. University of Houston - Victoria (Texas)
14. University of Michigan - Dearborn
15. University of North Alabama
16. Worcester State University (Massachusetts)

Within this cluster, seven institutions reside in the same geographic/accreditation region as APSU with Columbus State University, Jacksonville State University, The University of Tennessee-Martin, The University of Texas at Tyler, University of Houston-Clear Lake, University of Houston-Victoria, and the University of North Alabama sharing the same regional accreditation as APSU.

The cluster analysis indicated that cluster 4 was also close to APSU's home cluster, suggesting that further investigation should be conducted to determine if any of these institutions should be added to APSU's peer list or used to replace cluster 5 institutions outside the accreditation region. According to the analysis, those institutions within cluster 4 were:

1. Arkansas Tech University
2. Auburn University at Montgomery
3. Bowie State University (Maryland)
4. Eastern Illinois University

"With the cluster analysis indicating that cluster 5 was close to cluster 4, further investigations as to which institutions reside within this cluster should be done."

5. Framingham State University (Massachusetts)
6. Frostburg State University (Maryland)
7. McNeese State University (Louisiana)
8. Morehead State University (Kentucky)
9. Murray State University (Kentucky)
10. Radford University (Virginia)
11. Rhode Island College
12. Salisbury University (Maryland)
13. Sonoma State University (California)
14. Southern Oregon University
15. Texas A&M International University
16. The University of Tennessee-Chattanooga
17. The University of Nebraska at Kearney
18. University of North Carolina at Pembroke
19. University of Wisconsin-Platteville

Within this cluster, nine of the institutions share the same geographic/accreditation region. Therefore, some of these institutions could possibly be used to replace other institutions within the home cluster (cluster 5) that are not in the same accreditation region as APSU.

Table 3: APSU Values Compared to Primary and Secondary Clusters

Variables Used in Study	APSU Value	Primary Cluster Mean	Secondary Cluster Mean
Undergraduate enrollment for latest fall semester	9,513	6,166.56	7,431.29
Graduate enrollment for latest fall semester	831	1,411.25	1,021.11
FTE for latest academic year	12,144	9,666.25	9,374.74
Six-year graduation rate based on the IPEDS defined freshman cohort	37	42.21	46.68
Total operational revenues	68,464,626	60,032,510	78,339,202
Tuition and fee as percent of operational revenues	73.98	70.15	54.07
State appropriations as a percent of operational expenditures	55.39	58.49	53.03
Total expenditures	132,887,345	111,451,688	139,503,411
Instructional costs as a percent of expenditures	46.50	45.65	36.85
Endowment Assets per FTE	3,251.80	3,336.43	4,165.81
In-state tuition and fees on-campus	23,488	21,313.40	21,766.89
Out-of-state tuition and fees on-campus	38,728	31,279.47	31,527.74

The data listed in **Table 3** includes all of the twelve variables used in the study, APSU's value for each of these variables, and the mean values of each variable for cluster 5 (Primary), and cluster 4 (Secondary). From these data, it is clear that many of APSU's values more closely align with the means of cluster 5 than they do with cluster 4. The exceptions are the undergraduate enrollment where APSU is higher than cluster 5; graduate enrollment where APSU is closer to cluster 4, six-year graduation rate where APSU is lower than both means but closer to cluster 5; and both in- and out-of-state tuition where APSU is higher than both cluster means.

According to this study, The University of Tennessee-Martin is the only institution currently belonging to APSU's original IPEDS peer group that was present in cluster 5. The University of

Tennessee-Chattanooga is the only current Tennessee peer that was present in cluster 4.

The purpose of this study is to separate select institutions into similar clusters for use in determining comparator peers for APSU. The proximity of the cluster that the institutions are being compared should be considered as well as such factors as cost-of-living, non-traditional and international enrollment, location of major metropolitan areas close to the institution, and regional accrediting associations. These factors could significantly affect comparability within any model. Therefore, choosing institutions sharing the same regional accreditation as APSU was a major factor in DSIR's recommendation.

Recommendations

Based on the cluster analysis outcomes from this study, along with external factors such as cost-of-living and accreditation considerations, Decision Support and Institutional Research recommends 14 institutions that were within both clusters 5 and 4. The information within **Table 4** indicates the institution chosen, which cluster the institution was grouped, if the institution is in APSU's current list of peers, and if the institution was included within CUPA's list of peers for APSU.

All of these institutions are within ASPU's regional accreditation area and two of the Tennessee institutions were also in the original IPEDS peer list. Furthermore, all of the selected institutions were also included in the CUPA peer list. Therefore,

Table 4: Recommended Peer Institutions

Institution	Cluster No.	Current Peer	CUPA Peer
Columbus State University (Georgia)	5	No	Yes
Jacksonville State University (Alabama)	5	No	Yes
McNeese State University (Louisiana)	4	No	Yes
Morehead State University (Kentucky)	4	No	Yes
Murray State University (Kentucky)	4	No	Yes
Radford University (Virginia)	4	No	Yes
Texas A&M International	4	No	Yes
The University of Tennessee - Chattanooga	4	Yes	Yes
The University of Tennessee - Martin	5	Yes	Yes
The University of Texas at Tyler	5	No	Yes
University of Houston - Clear Lake (Texas)	5	No	Yes
University of Houston - Victoria (Texas)	5	No	Yes
University of North Alabama	5	No	Yes

this new list includes both institutions from the original IPEDS peer list and the CUPA peer list while only including public 4-year Masters (Larger Programs) with enrollments between 6,000 and 16,000 students.

Additional analysis can be performed by using different variables from IPEDS. This allows the institution to more closely align itself with institutional peers based on specific variables. Furthermore, the institution could use the IPEDS initial screening to include a peer group that is confined to only one region of the US, or it could add institutions with a higher level Carnegie Classification in order to identify potential aspirational peers.

It is important to note that subsequent iterations of the cluster analysis can include variables that were not pulled from the IPEDS database. Clearly, the number of total variables can impact the reliability of the model given the relatively small number of institutions, and parsimony is preferred. However, the model does allow administrators to choose variables that more closely align with the institution's role, scope, and mission in order to create a more meaningful institutional peer group.

While cluster analysis is clearly an exploratory data analysis technique for classifying and organizing institutions into meaningful groups, the results of such analyses are not definitive and should be reviewed with other quantitative and qualitative criteria. These methods, however, can save time and resources as institutions seek to find peer institutions to match their benchmarking needs.

"Clearly, the number of total variables used can impact the reliability of the model given the relatively small number of institution, and parsimony is preferred. However, the model does allow administrators to choose variables that more closely align with the institution's role, scope, and mission..."

Appendix I

Standardizing all quantifiable variables used in the analysis

Many researchers have noted the importance of standardizing variables for multivariate analysis. Otherwise, variables measured at different scales may not contribute equally to the analysis. This practice holds true for cluster analysis. Because of the sensitivity of most cluster models, raw values used for the variables may significantly alter the outcomes.

For example, in selecting peer institutions, a variable that ranges between $5 million and $10 million will influence significantly and have more weight in the analysis than a variable that ranges between 20 and 50. Therefore, transforming the data to comparable scales can prevent this problem. Typical data standardization procedures equalize the range and/or data variability. In the case of this study, variable values were standardized using z-scores with a mean of zero and a standard deviation of 1.

The z-score is a very useful statistic because it allows researchers to calculate the probability of a score occurring within the normal distribution and it enables researchers to compare two scores from different normal distributions. The standard score converts scores in a normal distribution to z-scores using the following formula:

$$z = \frac{x - \overline{x}}{S}$$

where x represents an individual score or observation in a set of scores, $\overline{x}$ represents the average of all individual scores or observations, and S represents the standard deviation of the scores or observations.

The z-score is synonymous to the standard deviation. A z-score of 2 is essentially 2 standard deviations above and below the mean. A z-score of 1.5 is 1.5 standard deviations above and below the mean. A z-score of 0 is equal to the mean of the distribution.

Z-scores exist on both sides of the mean. For example, 1 standard deviation below the mean is a z-score of -1 and a z-score of 2.2 can be 2.2 standard deviations above the mean. A z-score of -3 is 3 standard deviations below the mean. Put another way, the standard deviation and z-scores are just the average distance that individual values are from the mean.

Appendix II

Running the Cluster Analysis Procedure Using FASTCLUS within SAS

While there are numerous ways in which clusters may be formed, hierarchical clustering is one of the most straightforward methods. It can be either agglomerative or divisive. Agglomerative hierarchical clustering begins with each institution being a cluster unto itself. At successive steps, similar clusters are merged. The algorithm ends with all institutions in one, but useless, cluster. Divisive clustering starts with all institutions in one cluster and ends with each institution in its own cluster which, again, is not helpful. To find a good cluster solution, the researcher must look at the characteristics of the clusters at successive steps and decide when an interpretable solution is found that has a reasonable number of fairly homogeneous clusters.

This study used PROC FASTCLUS within SAS to determine the clusters. While the FASTCLUS procedure is intended for larger data sets, it can be used with smaller, although it can be sensitive to the order of the observations within the data set. This issue can be negated by standardizing the variables. PROC FASTCLUS also uses algorithms that place a large influence on variables with larger variance. Again, standardizing the variables before performing the analysis is highly recommended.

Table 1: Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	4.561	2.565	0.3801	.03801
2	1.996	0.559	0.1663	0.5464
3	1.437	0.021	0.1197	0.6661
4	1.256	0.296	0.1030	0.7691
5	.9391	0.196	0.0783	0.8474
6	.7433	0.254	0.0619	0.9093
7	.4890	0.186	0.0408	.09501
8	.3105	0.174	0.0259	0.9759
9	.1361	0.020	0.0113	0.9873
10	.1157	0.090	0.0096	0.9969
11	.0253	0.013	0.0021	0.9990
12	.0117		0.0010	1.0000

PROC FASTCLUS performs a disjoint cluster analysis on the basis of distances computed from one or more quantitative variables. The observations are divided into clusters so that every observation belongs to one cluster. By default, PROC FASTCLUS uses Euclidean distances, so the cluster centers are based on least squares estimation. The cluster centers are the means of the observations assigned to each cluster when the algorithm is run to complete convergence. PROC FASTCLUS is designed to find good clusters, not the best possible clusters, with only two or three

iterations of the data set and changing the number of clusters requested. This procedure can be effective in detecting outliers which appear as clusters with only one institution.

To run the analysis a two-step process was used to determine the number of possible clusters. This process used the CLUSTER procedure within SAS in order to examine eigenvalues, differences, and proportions. According to **Table 1**, a large difference exists between the first (4.561) and second (1.996) eigenvalues, proportions go from .3801 to .1663, with the cumulative proportion for the second eigenvalue equal to .5464. While this seems significant, a total of 73 institutions within only two clusters would be considerably under specified and the cumulative proportion indicates more clusters could be formed.

Upon further examination of the table, there exists a moderate change from eigenvalues eight (.3105) and nine (.1361), proportions go from .0259 to .0113, with the cumulative proportion for the ninth eigenvalue equal to .987 which is not much different from the cumulative percentage of .976 at eigenvalue eight. Further investigation revealed that clusters greater than eight would not contribute significantly to the model. Therefore, eight clusters were examined with results from PROC FASTCLUS.

Running the FASTCLUS procedure on eight clusters generated a significant Pseudo F Statistic of 12.83 and an observed overall R-Squared value of .58. The multivariate statistics and F approximations were then computed to test the fit of the model and the Wilks' Lambda, Pillai's Trace, Hotelling-Lawley Trace, and Roy's Greatest Root all confirmed that the model was significant with eight clusters.

Appendix III

An Analysis of CUPA Peer Institutions

As part of APSU's process of analyzing on-campus salaries, the institution developed a comparator group of public institutions in 2002. In 2010, there was a revised compensation study and the peer group was slightly modified to its current 68 institutions. The comparator group has been used by CUPA for regularly-scheduled salary studies.

In examining the CUPA peer list for possible inclusion as APSU's office peer comparator list, it was noted that some of the institutions were not aligned with APSU's role, scope, and mission.

Specifically, out of the 68 institutions chosen by CUPA, 19 had the Carnegie Classification of Doctoral/Research, 16 institutions had enrollments over 16,000 students, seven were Historically Black Colleges and Universities (HBCU), and seven had enrollments under 6,000. Below are the institutions used within CUPA salary studies:

1. Alabama Agricultural and Mechanical University
2. Appalachian State University
3. Arkansas State University
4. Armstrong State University
5. Auburn University at Montgomery
6. Bowie State University
7. College of Charleston
8. Columbus State University
9. Delta State University
10. East Carolina University
11. Eastern Kentucky University
12. East Tennessee State University
13. Fayetteville State University
14. Frostburg State University
15. Georgia College & State University
16. Georgia Southern University
17. Grambling State University
18. Jackson State University
19. Jacksonville State University
20. James Madison University
21. Marshall University
22. McNeese State University
23. Middle Tennessee State University
24. Morehead State University
25. Morgan State University – Doctoral Moderate Research 8,000
26. Murray State University
27. North Carolina Agricultural and Technical State University
28. North Carolina Central University
29. Northeastern State University
30. Northern Kentucky University
31. Northwestern State University
32. Prairie View A & M University

33. Radford University
34. Salisbury University
35. Sam Houston State University
36. Southeastern Louisiana University
37. Stephen F. Austin State University
38. Tarleton State University
39. Tennessee State University
40. Tennessee Technological University
41. Texas A&M International University
42. Texas A&M University - Corpus Christi
43. Texas A&M University - Kingsville
44. Texas State University
45. The University of Memphis
46. The University of Texas At El Paso
47. Towson University
48. Troy University
49. University of Central Arkansas
50. University of Central Oklahoma
51. University of Houston - Clear Lake
52. University of Houston – Victoria
53. University of North Alabama
54. University of North Carolina at Charlotte
55. University of North Carolina at Pembroke
56. University of North Carolina - Wilmington
57. University of North Georgia
58. University of South Alabama
59. University of Tennessee at Chattanooga
60. University of Tennessee at Martin
61. University of Texas at Tyler
62. University of West Alabama
63. University of West Florida
64. Valdosta State University
65. Western Carolina University
66. Western Kentucky University
67. West Texas A & M University
68. Winthrop University

Appendix IV

Institutions Used in the Cluster Analysis Study

Institutions used in the study included all public 4-year institutions with a Carnegie Classification of Masters - Larger Programs, enrollment between 6,000 and 16,000, and were not classified as HBCU, Land Grant, or Tribal institutions. The following institutions were used in the study:

1. Appalachian State University
2. Arkansas State University-Main Campus
3. Arkansas Tech University
4. Auburn University at Montgomery
5. Austin Peay State University
6. Bowie State University
7. Bridgewater State University
8. California State University-Chico
9. California State University-Dominguez Hills
10. Central Washington University
11. Coastal Carolina University
12. College of Charleston
13. Columbus State University
14. Eastern Illinois University
15. Eastern Kentucky University
16. Emporia State University
17. Fitchburg State University
18. Framingham State University
19. Frostburg State University
20. Georgia College & State University
21. Indiana University-Southeast
22. Jacksonville State University
23. James Madison University
24. Marshall University
25. McNeese State University
26. Morehead State University
27. Murray State University
28. North Carolina Central University
29. Northeastern Illinois University
30. Northeastern State University
31. Radford University
32. Rhode Island College
33. Saint Cloud State University
34. Salem State University
35. Salisbury University
36. Sonoma State University
37. Southeast Missouri State University
38. Southeastern Louisiana University
39. Southern Connecticut State University
40. Southern Oregon University
41. Southern Utah University
42. State University of New York at New Paltz
43. Stephen F Austin State University
44. SUNY College at Oswego
45. SUNY College at Plattsburgh

46. SUNY Cortland
47. Tarleton State University
48. Texas A & M International University
49. The College of New Jersey
50. The University of Tennessee-Chattanooga
51. The University of Tennessee-Martin
52. The University of Texas at Tyler
53. Towson University
54. Troy University
55. University at Buffalo
56. University of Central Arkansas
57. University of Central Missouri
58. University of Central Oklahoma
59. University of Houston-Clear Lake
60. University of Houston-Victoria
61. University of Michigan-Dearborn
62. University of Nebraska at Kearney
63. University of North Alabama
64. University of North Carolina at Pembroke
65. University of North Carolina - Wilmington
66. University of North Georgia
67. University of Wisconsin-Platteville
68. Western Carolina University
69. Western Illinois University
70. Western Kentucky University
71. Western Washington University
72. Winthrop University
73. Worcester State University

References

McLaughlin, G.W. and McLaughlin, J.S. (2007). The information mosaic: strategic decision making for universities and colleges. AGB: Washington, DC.

Teeter, D. J and Brinkman, P.T. (1987). Peer institutional studies/institutional comparisons. Primer for Institutional Research, J. Muffo and G. McLaughlin (eds), Association for Institutional Research: Tallahassee.

Terenzini, P. T., Hartmark, L., Lorang, W. G., & Shirley, R. C. (1980). A conceptual and methodological approach to the identification of peer institutions. Research in Higher Education, 12, 347-364.