Σtatistics at KΣU

Every Georgian Counts

Final Report on Sampling and Modeling

Prepared by Jennifer Lewis Priestley, Ph.D.

May 2008

Acknowledgements

This report on the status of the Every Georgian Counts project to better meet the needs of Georgia's homeless, represents the hard work and commitment of many individuals across several organizations, including, Don Watt, John Bassett, and Jane Massey from the Georgia Department of Community Affairs, Carol Pierannunzi and Christy Storey of the A.L. Burruss Institute of Public Service at Kennesaw State University, Vic Kane of the Department of Mathematics and Statistics and students Shan Muthersbaugh and Terry Jorgensen of Kennesaw State University.

Restatement of Objectives

The document ***Request for Proposal - Sampling Methodology for Count of Homelessness in Georgia***, required three primary activities for this initiative:

1. Develop a Methodology to Count Georgia's Homeless Population
2. Test this Methodology
3. Provide Recommendations for Improvement of HMIS

The results of the first two activities are reported in the present document. The results of third activity will reported under separate cover from A.L. Burruss Institute of Public Service.

Results of Activity 1 - Develop Methodology to Count Georgia's Homeless Population

In an effort to better meet the needs of Georgia's homeless, the Georgia Department of Community Affairs (DCA) sought to determine the number and characteristics of homeless persons in the state. By obtaining a reliable census of the homeless population, the state of Georgia would be better able to more effectively and efficiently utilize its resources to meet the needs of the homeless and to analyze the use and effectiveness of current assistance.

DCA was very clear in the original RFP document that the proposed methodology to obtain counts of the homeless population could not include a complete physical census of the state - *The proposed methodology should not be based on a complete statewide point-in-time count of unsheltered homeless. A physical count that includes field teams covering all census blocks in the state would be prohibitively difficult and expensive*.

Instead, DCA required a statistically valid sample of counties (the chosen spatial unit for analysis) for the purposes of conducting a physical count of homeless individuals in January 2008.

One approach to this requirement could have been the recommendation of a simple random sample taken from the 159 counties in Georgia. A simple random sample would have entailed randomly selecting some number of counties – say 10 – from the total of 159, where every county has a 10/159 probability of selection (approximately 6.3%).

The problem with this simplistic approach is two-fold:

1. Simple random sampling is subject to sampling error - "the luck of the draw". Without any pre-specified conditions, a random selection of counties could result in a selection of very similar counties where certain demographic or economic factors are not accommodated. For example, a simple random sample of 10 counties could yield 10 rural counties in the southern part of the state with very small populations, with a high percentage of the local economies based in agriculture. Alternatively, a sample of 10 counties could result in 10 counties from the Atlanta metropolitan area. Although both of these results are possible using a simple random sampling methodology, neither of these samples would be considered to be representative of the state as a whole.

A random sample based on the most "convenient" counties (selection based upon relationships or facilities) would not be appropriate either. This non-probabilistic approach would be analogous to "mall intercepts" or, in statistical terms "convenience sampling". While directionally interesting, the results are not statistically valid and should not be used for inference.

2. Homelessness is, by definition, heavily correlated with the size of the population – if there are more people, one would expect more homeless individuals. For example, it would be not be a surprise to learn that there are more homeless people in Fulton county (population 915,623[1]) than in Lowndes county (population 92,115[1]). However, what might be somewhat of a surprise is that there are <u>proportionately</u> more homeless in Lowndes county (.27% of the population[2]) than in Fulton county (.23% of the population). A random sample based on population would not accommodate this subtle, but very important, difference.

In an effort to address these issues, we developed a multi-stage sampling methodology, which incorporated the demographic and economic differences of the counties, while incorporating population and proportional considerations. The resulting sampling methodology utilized a combination of cluster analysis and proportional simple random sampling.

The first stage of the sampling methodology utilized cluster analysis. Cluster analysis is an unsupervised (no dependent variable is assigned)

---

[1] Source: 2007 Georgia County Guide. 2005 Population Estimate.

[2] Source: Unsheltered Homeless Count as provided by DCA, divided by the 2005 Population Estimate.

multivariate technique, which assigns observations into similar groupings or clusters, based upon the observational values across several variables of interest. In the present context, the 159 counties in GA represented 159 observations. From these 159 counties, 9 clusters were generated, based upon 300 variables taken from the 2000 U.S. Census[3]. Cluster size ranged from a low of 5 counties to a high of 35 counties. See Appendix 1 for cluster assignments[4].

These clusters are internally similar and externally different; counties within a particular cluster would be expected to demonstrate similar demographic and economic characteristics, while counties outside of a particular cluster would be expected to demonstrate very different characteristics. As a result, a county pulled from cluster 6 for example, would be expected to be similar to any other county in cluster 6 – they would be, in effect, interchangeable in everything except population size. It is for this reason that cluster analysis is heavily used by organizations, which collect data from individuals or groups spread across large areas, such as the U.S. Census Bureau.

The majority of the economic and demographic data in the present study was taken from the 2000 U.S. Census. The initial census data files included over 16,000 variables. These variables were condensed down to approximately 300 variables, which were then scaled to a common set of units - % of the county population. For example – If the original variable was "Total Population of Hispanics", this variable was divided by the total population for the county and converted to "Percent of County Population Identified as Hispanic". This was an important step, as the cluster analysis technique uses a Euclidean Distance algorithm to assign observations. These distances are calculated using the original units of the variable in question. As a result, variables scaled differently would generate wildly different distances. For this reason, all variables were expressed in common units of percent of population.

Because total population was not used as a basis for cluster generation, some county assignments may appear initially to be counter-intuitive. For example, Fulton County was found to be more similar to counties like Bartow and Houston, than counties like Cobb and DeKalb.

A few of the key variables used in the cluster analysis methodology and how they differ by cluster are presented below.

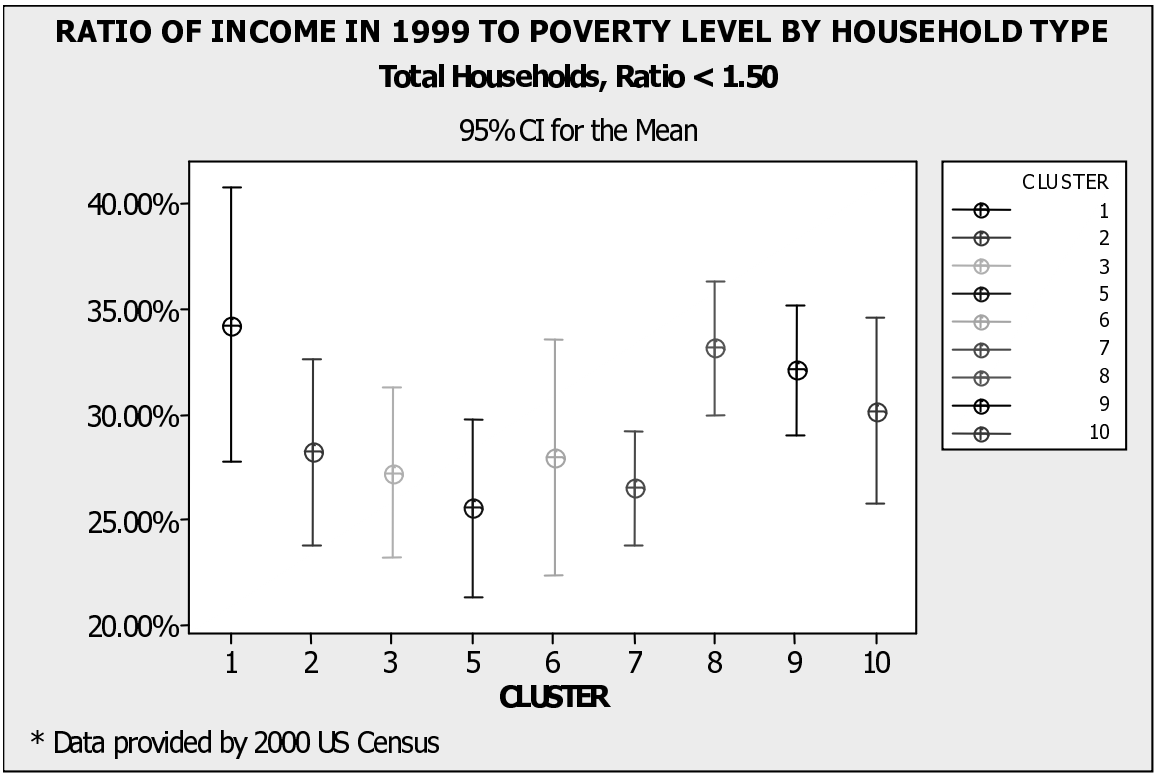
---

[3] Originally, 10 clusters were developed. However, one cluster – cluster 4 – was eventually dropped, and the counties were re-assigned.

[4] A color map, depicting the 9 clusters is available from Jennifer Priestley at jpriestl@kennesaw.edu.

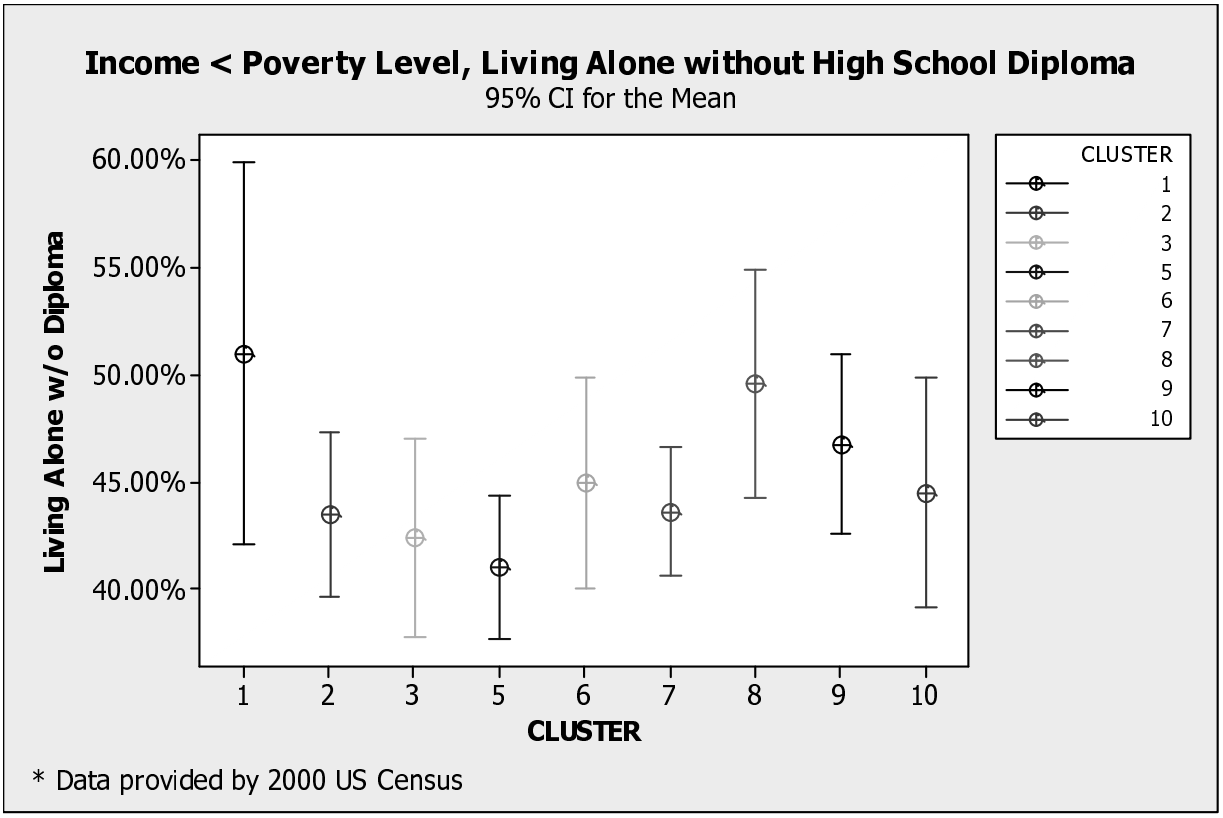
In Graph 1, the differences among the clusters for the percentage of households with gross income less than 150% of the poverty level can be seen. The individual vertical bars indicate the range of the 95% confidence interval, with the cluster mean represented as the circle in the middle of the line. The less overlap among the lines, the greater the differences by cluster.

Graph 1.



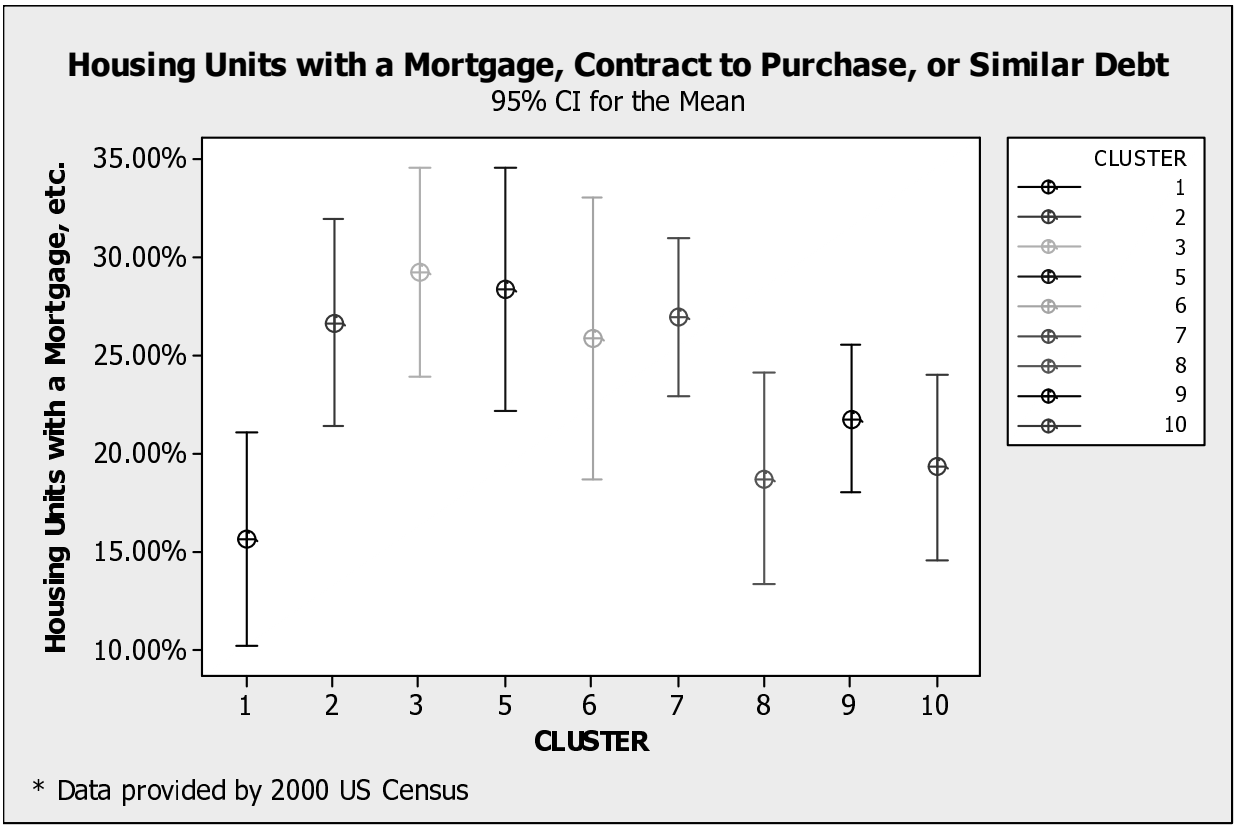
Graph 2, is conceptually similar, representing the differences among the clusters for the percentage of individuals without a high school diploma and a household income of less than 150% of the poverty level.

Graph 2.



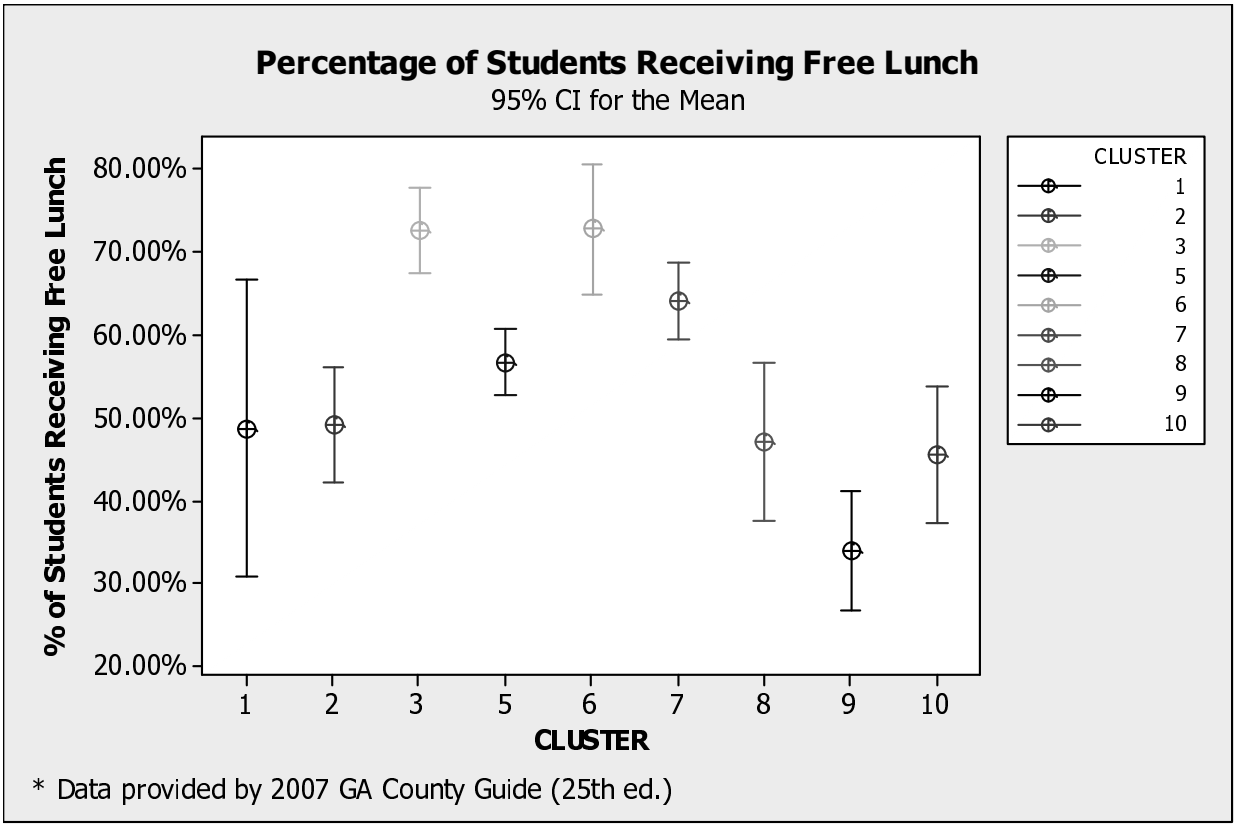
Graph three provides the differences among the clusters for the percentage of total housing units with a mortgage contract or similar contract to purchase.

Graph 3.



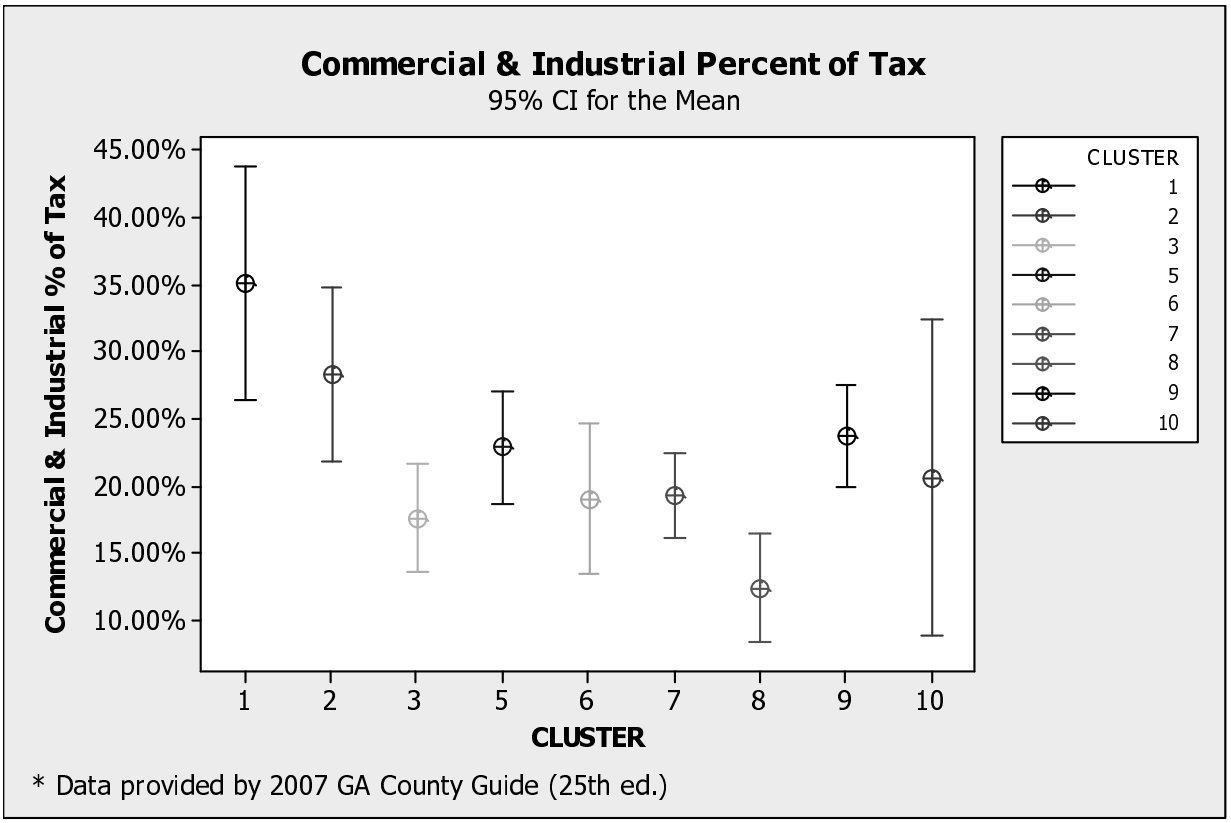
Graph 4, depicts the differences among the clusters, based upon the percentage of students qualifying for a free lunch. This graph (and graphs 5 and 6) are particularly interesting because this information was NOT part of the original clustering process, yet the differences among the clusters is clear – providing further evidence for the existence of natural groupings of counties within the state.

Graph 4.


Percentage of Students Receiving Free Lunch
95% CI for the Mean
% of Students Receiving Free Lunch
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
1
2
3
5
6
7
8
9
10
CLUSTER
CLUSTER
1
2
3
5
6
7
8
9
10
* Data provided by 2007 GA County Guide (25th ed.)

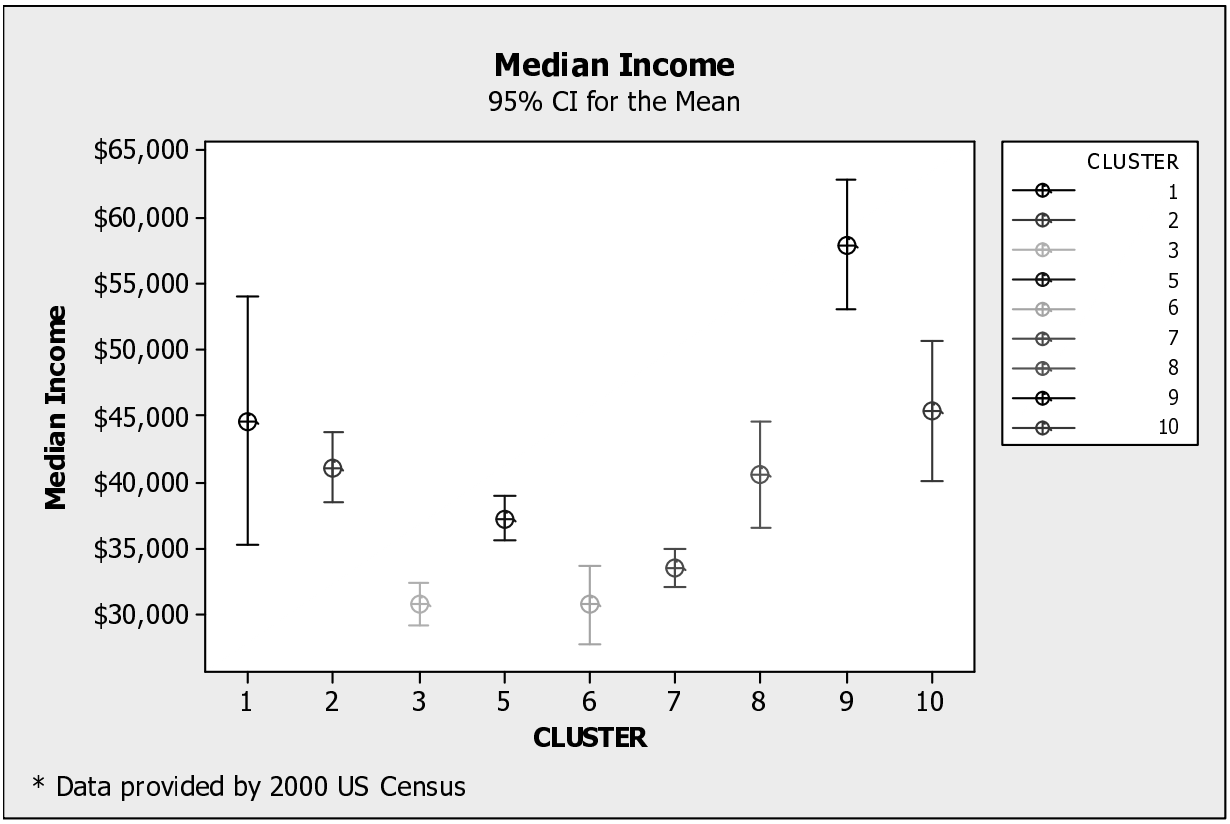
Graph 5 displays the differences among the clusters of the percentage of the gross tax digest, from commercial and industrial land.

Graph 5.



Finally, Graph 6 is a visualization of the differences in median income across the clusters. As noted above, this variable was NOT included in the original clustering methodology. As a result, the large differences across the clusters provide additional evidence to the existence of county clusters.

Graph 6.



As stated above, the clusters are internally similar and externally different from other clusters[5]. As a result, there is an inherent assumption, that any county selected within a cluster, is reasonably similar in economic and demographic characteristics to every other county within that cluster. Therefore, once the county assignments to the clusters were complete, sampling within each cluster could begin. Since every one of the clusters was to be sampled, this process addressed the first issue stated above – ensuring representation of different economic and demographic factors.

[5] ANOVA and Discriminant Analysis were used to provide initial validation of clusters. The results of this analysis are available upon request from the author.

Results of Activity 2 – Test this Methodology

The ratio-estimates of the homeless population, based upon the nine clusters outlined above, the resulting estimated counts and the actual values from the January 2008 count are included in Table 1.

Table 1: Results from January 2008 count

County Name	2000 Population	Cluster Assignment	Estimated Number of Homeless Based Upon Cluster Assignment	Regression Model Prediction of Homeless	January 2008 Count*	Variance from Ratio Prediction	Variance from Model Prediction
Chatham County	248,469	1	158	235	257	62.66%	9.36%
Fulton County	992,137	1	1862	1353	***1960***	5.26%	44.86%
Houston County	131,016	1	115	53	12	-89.57%	-77.36%
Richmond County	197,372	1	31	192	54	74.19%	-71.88%
						5.40%	**24.55%**
Bibb County	153,887	2	90	223	***171***	90.00%	-23.32%
Carroll County	87,268	2	51	106	52	1.96%	-50.94%
Dougherty County	96,065	2	56	144	39	-30.36%	-72.92%
Lowndes County	92,115	2	246	154	***32***	-86.99%	-79.22%
Muscogee County	186,291	2	108	112	352	225.93%	214.29%
						103.14%	**6.60%**
Burke County	22,243	3		53	47		-11.32%
							-11.32%
Clarke County	101,489	5	130	176	159	22.31%	-9.66%
Madison County	25,730	5	55	43	15	-72.73%	-65.12%
						-5.95%	**-20.55%**
Pierce County	15,636	6		37	0		
Mitchell County	23,932	7		34	2		-94.12%
Glynn County	67,568	8		122	31		-74.59%
Barrow County	46,144	9	19	43	7	-63.16%	-83.72%
Clayton County	236,517	9	140	73	56	-60.00%	-23.29%
Cobb County	607,751	9	251	297	***330***	31.47%	11.11%
DeKalb County	665,865	9	157	105	***155***	-1.27%	47.62%
Gwinnett County	588,448	9	243	97	127	-47.74%	30.93%
Walton County	60,687	9	25	43	3	-88.00%	-93.02%
						-21.20%	**3.04%**
Gordon County	44,104	10	59	60	27	-54.24%	-55.00%
Whitfield County	83,525	10	112	123	180	60.71%	46.34%
						21.05%	**13.11%**
GEORGIA			3908	3878	4068	**-0.77%**	**4.09%**

* Numbers in italics were collected in 2007

The 2007 and 2008 counts, which were taken during the last week in January, utilized the sampling methodology outlined above. The results of the counts – in the column titled "January 2008 Count" are compared to two sets of predictions.

The first prediction was developed using cluster-based ratios of homeless to the general population. This prediction can be seen above in the column titled "Estimated Number of Homeless Based Upon Cluster Assignment". The variance between these predictions and the January 2008 Count values can be seen in the column titled "Variance from Ratio Prediction". In this column, a negative value indicates that the prediction was greater than the "actual". Because of the challenges surrounding the accurate enumeration of the homeless population, it would be assumed that a negative value in this column would be preferred to a

positive value – indicating that the prediction was higher than the actual count.

Although most of the variances in this column are negative, several important (high population) county predictions were positive. Specifically, Chatham County (Savannah), Fulton County (Atlanta), Muscogee County (Columbus), Bibb County (Macon) and Cobb County (North Suburban Atlanta) all had cluster-based predictions lower than the January 2008 Count. Further investigation of these values provided an important insight. The homeless count value is based upon the cluster-based **ratio** of homelessness multiplied by the county population. All of the population values come from the 2000 US Census. Although not "official", the US Census does provide periodic updates at the county level. The most recent updates (2007) from the US Census are included in Table 2 below, with the revised predictions.

Table 2: Revised Ratio Predictions Using 2007 US Census Update

County Name	2007 Population Estimate from US Census	Cluster Assignment	Estimated Number of Homeless Based Upon Cluster Assignment	January 2008 Count*	Variance from Cluster Based Prediction
Chatham County	248,469	1	169	257	51.91%
Fulton County	992,137	1	2,264	***1960***	-13.42%
Houston County	131,016	1	136	12	-91.19%
Richmond County	197,372	1	31	54	76.31%
			2,600	2,283	**-12.19%**
Bibb County	154,709	2	90	***171***	90.00%
Carroll County	111,954	2	51	52	1.96%
Dougherty County	95,693	2	56	39	-30.36%
Lowndes County	101,790	2	246	***32***	-86.99%
Muscogee County	187,046	2	109	352	222.94%
			552	646	**17.03%**
Burke County	22,754	3		47	
Clarke County	114,063	5	130	159	22.31%
Madison County	28,012	5	55	15	-72.73%
			185	174	**-5.95%**
Pierce County	17,881	6		0	
Mitchell County	24,139	7		2	
Glynn County	74,932	8		31	
Barrow County	67,139	9	28	7	-74.85%
Clayton County	272,217	9	161	56	-65.25%
Cobb County	691,905	9	287	***330***	15.04%
DeKalb County	737,093	9	175	***155***	-11.38%
Gwinnett County	776,380	9	322	127	-60.55%
Walton County	83,144	9	34	3	-91.30%
			1,007	678	**-32.68%**
Gordon County	52,044	10	59	27	-54.24%
Whitfield County	93,379	10	112	180	60.71%
			171	207	**21.05%**
GEORGIA			**4,515**	**3,988**	**-11.67%**

** Numbers in italics were collected in 2007*

As can be seen in Table 2, the prediction for the major population center of Fulton County, which accounts for over 10% of the entire population of the state, is 2,264 homeless, where the January 2007 count was 1,960. Although the variances for the other four counties are still positive, the predictions are closer to the actual count values. In addition, the overall variance changed from a value of negative .77% using the 2000 Census data to a negative value of 11.67% using the 2007 Census updates for the major population centers.

This simple approach of taking the past homeless data for a cluster and dividing by the total population for a period in time and then using that ratio against a future population estimate, produced surprisingly strong results. The primary disadvantage to estimating the homeless population using this approach is that there is no "model" – there are no predictors or explanatory variables, as is the case with regression modeling. While there

is economic and demographic data used in the creation of the clusters, this kind of information is not used directly to predict or explain the issue of homelessness.

The second set of predictions from Table 1 – The Regression Model Prediction of Homelessness – were similarly compared to the actual values from the January 2008 count. The variances associated with these predictions suffered from the same issue of using the 2000 US Census population figures. As with the previous predictions, when the population figures were updated using the US Census 2007 estimates, the variances improved – from positive 4.09% to negative 8.69%.

Table 3: Revised Model Predictions Using 2007 US Census Update

County Name	2007 Population Estimate from US Census	Cluster Assignment	Regression Model Prediction of Homeless	January 2008 Count*	Variance from Model Prediction
Chatham County	248,469	1	252	257	1.98%
Fulton County	992,137	1	1,646	***1960***	19.07%
Houston County	131,016	1	63	12	-80.86%
Richmond County	197,372	1	190	54	-71.55%
			2,151	2,283	**6.16%**
Bibb County	154,709	2	225	***171***	-24.03%
Carroll County	111,954	2	137	52	-61.97%
Dougherty County	95,693	2	144	39	-72.92%
Lowndes County	101,790	2	171	***32***	-81.24%
Muscogee County	187,046	2	112	352	214.29%
			788	646	**-18.07%**
Burke County	22,754	3	55	47	-13.77%
Clarke County	114,063	5	198	159	-19.85%
Madison County	28,012	5	47	15	-68.40%
			246	174	**-29.22%**
Pierce County	17,881	6	43	0	
Mitchell County	24,139	7	35	2	-94.31%
Glynn County	74,932	8	136	31	-77.21%
Barrow County	67,139	9	64	7	-89.01%
Clayton County	272,217	9	85	56	-34.14%
Cobb County	691,905	9	339	***330***	-2.54%
DeKalb County	737,093	9	117	***155***	32.83%
Gwinnett County	776,380	9	129	127	-1.73%
Walton County	83,144	9	59	3	-94.93%
			792	678	**-14.44%**
Gordon County	52,044	10	71	27	-61.99%
Whitfield County	93,379	10	138	180	30.11%
			209	207	**-1.14%**
GEORGIA			**4,455**	**4,068**	**-8.69%**

** Numbers in italics were collected in 2007*

Although the overall variance for these predictions was lower than the overall variance for the cluster-based ratios, it should be noted that two of the largest counties in the state – Fulton and DeKalb had positive

variances. However, Cobb and Bibb counties, which had positive variances in Table 2, have negative variances in Table 3.

Once the "actual" counts were available, the Multiple Regression models from the interim report to the DCA were not just "rerun", but actually redeveloped using a combination of variables from the 2000 US Census and the 2007 Georgia County Guide. A stepwise selection methodology was used.

The revised model included 7 variables:

Variable Name	Parameter Estimate	T-Value	p-value
Intercept	0.00051273	.93	0.3523
Per Capita Income in 1999	.00000007	2.98	0.0034
Percent of HHs with Income <1.5 of the Poverty Level	0.01422	3.37	0.001
Percent of Housing Units for Rental, Not Occupied	.00342	1.48	.143
Percent of Housing Units with a Mortgage or other contract to purchase	-0.01525	-7.17	<.0001
Percent of Housing Units with no Mortgage contract	0.01512	5.67	<.0001
Percent of Housing Units Lacking Complete Plumbing Facilities	-0.12188	-2.72	0.0073

Using these variables, the percent of homeless, by county can be made. These percentages (ratios) can then be applied against the population values to generate homeless predictions by county. The full listing of predictions, by county, can be seen in Appendix 3.

It is worth noting that the overall estimate of unsheltered homeless individuals in Georgia is 12,058. Based on a 2007 total population of 9,544,750, this prediction would indicate that approximately .1263% of individuals in the state of Georgia are unsheltered homeless. In the February 2007Annual Homeless Assessment Report to Congress, the authors indicated that the number of unsheltered homeless was 338,781 (page iii in the Executive Summary). Based upon a U.S. population of 296,507,061 in 2005 (when the unsheltered data was collected), this would indicate that the U.S. rate of unsheltered homeless is .1143%. The

prediction generated using the current model for the State of Georgia is higher than the HUD prediction by less than .012%.

Issues and Limitations

The great statistician George Box is attributed with the quote "*Essentially, all models are wrong, but some are useful".*

The sampling methodology and prediction model provided in this document are not 100% accurate, but should be useful.

As with all analytical exercises, the present analysis has limitations. Specifically, there are two primary issues/limitations to which the reader should be alerted.

1. The clustering procedure was based upon the data from the 2000 U.S. Census. More than any other data source, the census data provided a vast, rich assortment of demographic and econometric variables, which are not readily available through any other source. This data also exists at many different levels of aggregation, including county and census tract. These points made this data particularly attractive for finding natural groupings. However, at a minimum, this data is eight years old. As populations migrate and economies change, the data from the 2000 Census becomes outdated. As a result, the clusters defined in the present report might be differently configured with the 2010 Census.

In addition, the 2000 U.S. Census data was used as the predictor variables, while the dependent variable estimations of unsheltered homeless were based on 2007/8 counts.

2. Anyone who has engaged in statistical modeling at any point would have expected to see an "R^2" value reported with the regression model. In brief, this value is the primary metric used to understand how well the independent variables explain the percentage variance (change) in the dependent variable. In this case, proportion of unsheltered homeless. This value was not discussed for the present model. Why?

Recall that the number of unsheltered homeless was only provided for 23 out of 159 counties. This was an insufficient number to engage in a modeling exercise. As a result, the proportion of homeless for the other counties was estimated from this small amount of data. The model was then developed from this estimation.

The eventual model did generate an adjusted R^2 value of 70%. This would indicate that the model can explain 70% of the change in the proportion of homelessness AS ESTIMATED. If the estimates were changed, the model and the R^2 value would change.

Another related point is unsheltered homeless individuals are difficult to count; the confidence in some of the "actual" numbers may be low. As a result, the accuracy of the predictions from the model becomes somewhat of a moving target. For example, if the "actual" count for a county is 100 but the model predicted 150 for the county, there is a possibility that, given the characteristics of the county, the count is an under representation of the actual homeless. It is for this reason that negative variances were preferred to positive variances.

These issues make the traditional R^2 metric less meaningful in the present study.

Bibliography

1. US Census Bureau. (2000). United States Census 2000. Retrieved April, 2006 from http://www.census.gov/main/www/cen2000.html.

2. US Department of Housing and Urban Development, Office of Community Planning and Development. (2007). *The Annual Homeless Assessment Report to Congress*. Washington D.C.

3. United Way of Metropolitan Atlanta. (2003). *Blueprint to End Homelessness in Atlanta in Ten Years*. Atlanta, GA: Mark O'Connell.

4. Georgia Department of Community Affairs. (2005) *State of Georgia Five Year Consolidated Plan*. Atlanta, GA: Mike Beatty.

5. Center for Agribusiness and Economic Development. (2007) *Georgia County Guide.*

6. Derr, D. and Nagle, G. (1977). Land transfers at the Rural/Urban Fringe: The New Jersey Experience. *American Society of Farm Managers and Rural Appraisers Journal*. (41), 31-35.

7. Baum, S., Haynes, M., van Gellecum, Y., and Hoon Han, J. (2007) Considering regional socio-economic outcomes in non-metropolitan Australia: A typology building approach. *Regional Science*. (56), 261-285.

Appendix 1: Cluster Assignments of Georgia Counties

County	2006 Population*	Cluster
Bartow County	89,229	1
Chatham County	238,410	1
Fulton County	915,623	1
Houston County	126,163	1
Richmond County	195,769	1
Bibb County	154,918	2
Butts County	21,045	2
Camden County	45,759	2
Carroll County	105,453	2
Catoosa County	60,813	2
Dougherty County	94,882	2
Effingham County	46,924	2
Floyd County	94,198	2
Habersham County	39,603	2
Jones County	26,836	2
Lowndes County	96,705	2
Muscogee County	185,271	2
Peach County	24,794	2
Spalding County	61,289	2
Troup County	9,474	2
Atkinson County	8,030	3
Bacon County	10,379	3
Barker County	16,055	3
Brooks County	16,327	3
Burke County	23,299	3
Calhoun County	5,972	3
Candler County	10,321	3
Chattooga County	26,570	3
Clinch County	6,996	3
Dodge County	19,574	3
Early County	12,056	3
Glascock County	2,705	3
Hancock County	9,643	3
Irwin County	10,093	3
Jenkins County	8,729	3
Johnson County	9,538	3
Lincoln County	8,207	3
Quitman	2,467	3
Randolph County	7,310	3

Schley County	4,122	3
Stewart County	4,882	3
Tattnall County	23,211	3
Telfair County	13,205	3
Terrell County	10,711	3
Turner County	10,299	3
Twiggs County	19,782	3
Washington County	20,118	3
Wheeler County	6,706	3
Wilcox County	8,721	3
Baldwin County	4,154	5
Banks County	45,230	5
Bulloch County	61,454	5
Clarke County	104,439	5
Cook County	16,366	5
Crawford County	12,874	5
Dade County	16,040	5
Haralson County	28,338	5
Heard County	11,346	5
Lanier County	7,553	5
Laurens County	46,896	5
Madison County	27,289	5
McDuffie County	21,743	5
Meriwether County	22,919	5
Murray County	40,812	5
Pickens County	28,442	5
Polk County	40,479	5
Pulaski County	9,737	5
Stephens County	25,060	5
Thomas County	44,692	5
Tift County	40,793	5
Toombs County	27,274	5
Thomaston-Upson County	10,315	5
Walker County	63,890	5
Appling County	17,954	6
Clay County	3,242	6
Emanuel County	22,108	6
Jeff Davis County	13,083	6
Jefferson County	16,926	6
Miller County	6,228	6
Montgomery County	8,909	6
Pierce County	17,119	6
Seminole County	9,226	6

Talbot County	6,709	6
Taylor County	8,887	6
Treutlen County	62,015	6
Warren County	6,101	6
Wilkinson County	10,143	6
Ben Hill County	17,316	7
Berrien County	16,708	7
Bleckley County	12,141	7
Brantley County	15,491	7
Charlton County	10,790	7
Chattahoochee County	14,679	7
Coffee County	39,674	7
Colquitt County	43,915	7
Crisp County	22,017	7
Decatur County	28,618	7
Dooly County	11,749	7
Echols County	4,253	7
Elbert County	20,799	7
Evans County	11,443	7
Fannin County	21,887	7
Franklin County	21,590	7
Grady County	24,466	7
Greene County	15,693	7
Hart County	24,036	7
Macon County	13,745	7
Marion County	7,244	7
McIntosh County	11,068	7
Mitchell County	23,791	7
Putnam County	19,829	7
Rabun County	16,087	7
Screven County	15,430	7
Sumter County	32,912	7
Taliaferro County	1,826	7
Towns County	6,753	7
Union County	27,679	7
Ware County	34,492	7
Wayne County	28,390	7
Webster	2,289	7
Wilkes County	10,457	7
Worth County	21,996	7
Dawson County	19,731	8
Gilmer County	27,335	8
Glynn County	71,874	8

Jasper County	13,147	8
Long County	11,083	8
Lumpkin County	24,324	8
Morgan County	17,492	8
Oglethorpe County	13,609	8
White County	24,055	8
Barrow County	59,954	9
Cherokee County	184,211	9
Clayton County	267,966	9
Cobb County	663,818	9
Columbia County	103,812	9
Coweta County	109,903	9
DeKalb County	677,959	9
Douglas County	112,760	9
Fayette County	104,248	9
Forsyth County	140,393	9
Gwinnett County	726,273	9
Hall County	165,771	9
Henry County	167,848	9
Lee County	31,099	9
Newton County	86,713	9
Oconee County	29,748	9
Paulding County	112,411	9
Rockdale County	78,545	9
Walton County	75,647	9
Bryan County	28,549	10
Gordon County	50,279	10
Harris County	27,779	10
Jackson County	52,292	10
Lamar County	16,378	10
Liberty County	57,544	10
Monroe County	23,785	10
Pike County	16,128	10
Whitfield County	90,889	10
Total for State	9,072,576	

*Source: 2007 Georgia County Guide

Appendix 2: Required Sample Size by Cluster

Cluster	Cluster 2005 Population*	Required Population Sample Size
1	1,565,194	229,645
2	1,067,964	70,700
3	336,028	87,372
5	758,135	63,179
6	208,650	87,372
7	651,253	87,372
8	222,650	87,372
9	3,899,079	291,384
10	363,623	71,527

*Source: 2007 Georgia County Guide

Appendix 3: Revised Homeless Predictions Using the January 2008 Count

COUNTY NAME	2007 US Census Estimate	CLUSTER ASSIGNMENT	Predicted Homeless Ratio	Revised Count Prediction
Bartow County	92,834	1	0.0935%	87
Chatham County	248,469	1	0.1014%	252
Fulton County	992,137	1	0.1659%	2,660
Houston County	131,016	1	0.0479%	63
Richmond County	197,372	1	0.0962%	190
Cluster Total	**1,661,828**			**3,251**
Bibb County	154,709	2	0.1455%	225
Butts County	23,759	2	0.0944%	22
Camden County	48,689	2	0.0933%	45
Carroll County	111,954	2	0.1221%	137
Catoosa County	62,241	2	0.0931%	58
Dougherty County	95,693	2	0.1505%	144
Effingham County	50,728	2	0.0843%	43
Floyd County	95,618	2	0.1143%	109
Habersham County	42,272	2	0.1225%	52
Jones County	27,229	2	0.0922%	25
Lowndes County	101,790	2	0.1676%	171
Muscogee County	187,046	2	0.0604%	421
Peach County	25,672	2	0.0908%	23
Spalding County	62,826	2	0.1143%	72
Troup County	9,270	2	0.1069%	10
Cluster Total	**1,099,496**			**1,557**
Atkinson County	8,223	3	0.3506%	29
Bacon County	10,507	3	0.3908%	41
Barker County	16,556	3	0.2575%	43
Brooks County	16,340	3	0.1987%	32
Burke County	22,754	3	0.2395%	55
Calhoun County	6,098	3	0.1592%	10
Candler County	10,550	3	0.2746%	29
Chattooga County	26,797	3	0.2107%	56
Clinch County	6,992	3	0.3560%	25
Dodge County	20,042	3	0.3325%	67
Early County	11,836	3	0.2756%	33
Glascock County	2,771	3	0.3999%	11
Hancock County	9,568	3	0.3352%	32
Irwin County	9,934	3	0.2492%	25
Jenkins County	8,595	3	0.3619%	31
Johnson County	9,533	3	0.2917%	28
Lincoln County	8,098	3	0.4322%	35
Quitman	2,666	3	0.7396%	20
Randolph County	7,294	3	0.2503%	18
Schley County	4,123	3	0.1507%	6
Stewart County	4,647	3	0.3707%	17
Tattnall County	23,179	3	0.2772%	64
Telfair County	13,366	3	0.2786%	37
Terrell County	10,260	3	0.1537%	16
Turner County	10,280	3	0.2431%	25
Twiggs County	20,968	3	0.1952%	41
Washington County	20,937	3	0.1745%	37
Wheeler County	6,830	3	0.3389%	23
Wilcox County	8,613	3	0.2837%	24
Cluster Total	**338,357**			**909**
Baldwin County	3,781	5	0.1570%	6
Banks County	46,057	5	0.2009%	93
Bulloch County	66,176	5	0.1596%	106
Clarke County	114,063	5	0.1739%	198
Cook County	16,432	5	0.1875%	31
Crawford County	12,483	5	0.1670%	21
Dade County	16,098	5	0.1864%	30
Haralson County	28,718	5	0.2211%	63
Heard County	11,387	5	0.2548%	29
Lanier County	7,947	5	0.2652%	21
Laurens County	47,520	5	0.2346%	112
Madison County	28,012	5	0.1695%	47
McDuffie County	21,551	5	0.2054%	44
Meriwether County	22,748	5	0.1496%	34
Murray County	40,664	5	0.2155%	88
Pickens County	30,488	5	0.1881%	57
Polk County	41,460	5	0.2108%	87
Pulaski County	9,843	5	0.2116%	21
Stephens County	25,268	5	0.2179%	55
Thomas County	45,237	5	0.1356%	61
Thomaston-Upson County	41,610	5	0.1471%	61
Tift County	27,820	5	0.1371%	38
Toombs County	10,894	5	0.2356%	26
Walker County	64,554	5	0.1676%	108
Cluster Total	**780,811**			**1,438**

COUNTY NAME	2007 US Census Estimate	CLUSTER ASSIGNMENT	Predicted Homeless Ratio	Revised Count Prediction
Appling County	17,946	6	0.4030%	72
Clay County	3,207	6	0.5597%	18
Emanuel County	22,469	6	0.3345%	75
Jeff Davis County	13,291	6	0.3219%	43
Jefferson County	16,454	6	0.1907%	31
Miller County	6,163	6	0.2992%	18
Montgomery County	9,060	6	0.2713%	25
Pierce County	17,881	6	0.2398%	43
Seminole County	9,081	6	0.3722%	34
Talbot County	6,607	6	0.2778%	18
Taylor County	8,738	6	0.3594%	31
Treutlen County	63,535	6	0.2973%	189
Warren County	5,908	6	0.2973%	18
Wilkinson County	10,064	6	0.2180%	22
Cluster Total	**210,404**			**637**
Ben Hill County	17,650	7	0.2185%	39
Berrien County	16,722	7	0.2899%	48
Bleckley County	12,306	7	0.1634%	20
Brantley County	15,440	7	0.3222%	50
Charlton County	10,609	7	0.2425%	26
Chattahoochee County	9,430	7	0.0489%	5
Coffee County	40,085	7	0.2723%	109
Colquitt County	44,814	7	0.2156%	97
Crisp County	22,125	7	0.2114%	47
Decatur County	28,544	7	0.1965%	56
Dooly County	11,592	7	0.2217%	26
Echols County	4,093	7	0.4849%	20
Elbert County	20,525	7	0.2173%	45
Evans County	11,505	7	0.2065%	24
Fannin County	22,580	7	0.4176%	94
Franklin County	21,793	7	0.2459%	54
Grady County	25,042	7	0.2492%	62
Greene County	15,662	7	0.3253%	51
Hart County	24,240	7	0.2654%	64
Macon County	13,524	7	0.1257%	17
Marion County	7,024	7	0.2418%	17
McIntosh County	11,420	7	0.3754%	43
Mitchell County	24,139	7	0.1456%	35
Putnam County	20,251	7	0.4158%	84
Rabun County	16,519	7	0.6758%	112
Screven County	15,037	7	0.3108%	47
Sumter County	32,532	7	0.0755%	25
Taliaferro County	1,884	7	0.3321%	6
Towns County	6,938	7	0.6008%	42
Union County	27,562	7	0.4409%	122
Ware County	35,831	7	0.2234%	80
Wayne County	29,046	7	0.2017%	59
Webster	2,245	7	0.3140%	7
Wilkes County	10,262	7	0.3096%	32
Worth County	21,285	7	0.2346%	50
Cluster Total	**650,256**			**1,711**

COUNTY NAME	2007 US Census Estimate	CLUSTER ASSIGNMENT	Predicted Homeless Ratio	Revised Count Prediction
Dawson County	21,484	8	0.0799%	17
Gilmer County	28,389	8	0.3688%	105
Glynn County	74,932	8	0.1815%	136
Jasper County	13,660	8	0.1398%	19
Long County	11,300	8	0.2741%	31
Lumpkin County	26,554	8	0.1302%	35
Morgan County	18,165	8	0.0943%	17
Oglethorpe County	13,963	8	0.1935%	27
White County	25,020	8	0.1973%	49
Cluster Total	**233,467**			**436**
Barrow County	67,139	9	0.0949%	64
Cherokee County	204,363	9	0.0571%	117
Clayton County	272,217	9	0.0312%	85
Cobb County	691,905	9	0.0489%	339
Columbia County	109,100	9	0.0457%	50
Coweta County	118,936	9	0.0498%	59
DeKalb County	737,093	9	0.0158%	117
Douglas County	124,495	9	0.0635%	79
Fayette County	106,144	9	0.0248%	26
Forsyth County	158,914	9	0.0888%	141
Gwinnett County	776,380	9	0.0166%	129
Hall County	180,175	9	0.0636%	115
Henry County	186,037	9	0.0288%	54
Lee County	33,050	9	0.0373%	12
Newton County	96,019	9	0.0509%	49
Oconee County	31,367	9	0.0388%	12
Paulding County	127,906	9	0.0656%	84
Rockdale County	82,052	9	0.0364%	30
Walton County	83,144	9	0.0712%	59
Cluster Total	**4,186,436**		**0.0457%**	**1,620**
Bryan County	30,132	10	0.1251%	38
Gordon County	52,044	10	0.1365%	71
Harris County	29,073	10	0.1704%	50
Jackson County	59,254	10	0.1336%	79
Lamar County	16,961	10	0.0937%	16
Liberty County	60,503	10	0.1192%	72
Monroe County	25,145	10	0.0849%	21
Pike County	17,204	10	0.0704%	12
Whitfield County	93,379	10	0.1482%	138
Cluster Total	**383,695**		**0.1296%**	**497**
GEORGIA	**9,544,750**		**0.1263%**	**12,058**