New Market Groupings Based on Food Consumption Patterns

Ronald B. Larson
Department of Marketing, Haworth College of Business, Western Michigan University, Kalamazoo, MI 49008–5430. E-mail: ron.larson@wmich.edu

ABSTRACT

Consumption of different food categories varies across the US. A cluster analysis of 52 markets based on food sales in 62 categories found 11 market groupings with similar consumption patterns. These new market groupings were compared with a cluster analysis based on data from 10 years earlier. Many patterns were similar but a few shifts in consumption patterns were also noted. Evidence suggested that some regional food preferences may have fragmented during the last 10 years. Researchers, managers, and policy makers should use market groupings based on fairly current data when they examine food demand trends across geographics, test marketing programs, identify opportunities in selected markets, forecast results for regional expansions, or evaluate regional initiative performance. [EconLit citations: L660, R220, Q130.] © 2004 Wiley Periodicals, Inc.

1. INTRODUCTION

Researchers have long recognized variations in customer characteristics and preferences across the geographies. Before the Revolutionary War, British geographers noted that the colonies exhibited well-defined sectional patterns in climate, economy, society, and demographics (Mood, 1951). Many factors can contribute to the creation and evolution of different geographic consumption patterns. For example, a type of French liqueur that is very popular in one town in England is produced in the area of France where local solders were stationed during World War II (Gerrie, 1987). Grigg (1999) argued that the 750 years the Romans occupied the Mediterranean region were critical to the Mediterranean diet's development. Major crops that were the basis for this diet were grown throughout the region in response to demand from Rome. Rapid economic growth since 1960 changed the traditional Mediterranean diet and reduced the similarity of consumption patterns across the region.

During the 1980s, U.S. firms became more interested in regional purchase patterns and some locations were targeted with extra, more localized marketing efforts (McKenna, 1992; Yeager, 1987). The popularity of regional analysis has grown. In a series of books, Michael Weiss illustrated many U.S. regional patterns. In his first book, Weiss (1988) documented how values and behaviors varied across the US and described the 40 PRIZM clusters using data from several sources. Weiss (1994) showed regional consumption patterns for many foods and profiled 211 consumer market areas using indexes for lifestyles,

Agribusiness, Vol. 20 (4) 417–432 (2004)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/agr.20027

leisure activities, media usage, and product purchases. Weiss (2000) illustrated more regional patterns and described each of the 62 PRIZM clusters. According to Weiss, the increase in the cluster number was necessary because the country was fragmenting into more segments.

Geographic patterns of food consumption are of interest to researchers, managers, and policy makers. Researchers often need to control for regional variations and may need to update their geographic groupings. Some food demand analysts have used the Bureau of Census regions, developed in 1910, to control for geographic variations in food preferences. If these regions do not reflect current preference patterns, they may add significant bias to the results. Larson (1998) showed that models with updated geographies developed using cluster analysis tended to explain food sales patterns better than models using the Bureau of the Census regions. Research that uses regional data may need new market groupings.

Managers can use regional analysis to test marketing programs, to identify areas with growth opportunities, and to track changes in those opportunities. Food manufacturers may vary trade deal programs or employee incentives across the country to learn which works the best. Retailers may test different pricing or merchandising systems to learn which should be implemented across the chain. Regional information can help balance test and control cells for these experiments. This data can also make it possible to identify growth opportunities. For example, Grewal, Levy, Mehrotra, and Sharma (1999) evaluated the performance of 59 stores in a retail chain. By disaggregating the group into three regions and benchmarking stores against their region, they increased the number of stores judged to be efficient from 14 to 30. With geographic information, managers could focus on the 29 stores with the most growth potential.

Once a firm has started with a regional segmentation initiative, Weinstein (1987) suggests that managers should request frequent updates. Linneman and Stanton (1991) recommend that managers should "always be collecting data." However, regularly buying and analyzing geographic data is expensive. If regional purchase patterns tend to be stable, analyses will be needed less frequently. Few studies have looked at the stability of regional patterns.

Like managers, policy makers have questions about how consumer needs vary and how they may be changing. For example, if the diets in a region have not been particularly nutritious, what popular foods (e.g., fruits or vegetables) might household members be willing to eat more of to improve their diets? What other geographies may be particularly responsive to the same educational program because they have similar food preferences? If consumption patterns are particularly dynamic, educational initiatives to improve diets may need to be revised regularly.

The primary objective of this research is to develop new market groupings based on food sales indices. These groupings and the cluster analysis technique used to develop them may be useful to researchers, managers, and policy makers. The geographic stability of food sales patterns in the US will also be explored by comparing the new grouping with one based on data that is 10 years older. Some evidence of regional food consumption variations is reviewed in the first section of this paper. Then, the sales data and the methodology used to group the data are introduced. Results are discussed in the third section. Next, these results are compared with an earlier market grouping described in a previous paper (Larson, 1998). Possible shifts in food consumption patterns during the 10-year difference in the data periods are highlighted. The study's implications for researchers, managers, and policy makers are summarized in the final section.

2. VARIATIONS ACROSS REGIONS

The consumption of many foods varies by region. Urbanski (2003) reported that wine purchases were highest in the West (dollar sales index of 160) and lowest in the Central region (index of 75). Puddings and dairy dessert sales were highest in the East (index of 139) and lowest in the South (index of 74). Even consumption patterns for ethnic groups such as African Americans had significant regional differences (Harris & Nowverl, 2000). Moving from regions to markets usually reveals a larger range of consumption patterns. For example, Larson (1998) showed that family flour purchases per household varied from nearly 2.5 times the national average in the San Antonio/Corpus Christi area and more than twice the national average in Memphis/Little Rock and El Paso/Albuquerque/Lubbock to about 40% below the national average in Philadelphia and Los Angeles.

Demand researchers often add regional dummy variables to cross-section analyses to account for geographic differences in tastes or variations in price and promotion sensitivity. These variations can exist within smaller areas such as a state. Mulhern, Williams, and Leone (1998) analyzed four years of scanner data from 35 specialty liquor stores in one state. They found that income levels and the ethnic mix around the stores influenced each product's price and promotion sensitivity. Store displays without any price reduction or feature advertising tend to generate larger sales increases in some markets. Leeds (1994) reported that sales in Milwaukee and Portland tended to more than double when products were displayed without other marketing support. In the Southeast, the sales gains were much smaller with the increases in Miami and Tampa averaging only 57%. Failing to account for important factors that have a regional distribution, either directly or with regional dummy variables, can introduce biases into a demand analysis.

Marketers have tried to take advantage of geographic differences in preferences. For example, food marketers tailor their products to regional tastes. Frito Lay started offering vinegar-flavored chips in the Northeast, mesquite-flavored chips in the Southwest, and sour cream-flavored chips in the Midwest (Hiam & Schewe, 1992). McDonald's adjusted their menus to regional food preferences by selling McLobster Rolls in Maine, McTeri Burgers in Hawaii, and McDeli Wraps in Michigan (Blank, 1998). Another way marketers target geographies is with spot advertising. In 1998, about $14 billion was spent on spot advertising. An industry study concluded marketers were using poor criteria (brand sales indices) to select markets to target (Cardona, 1998). Sales gains from spot advertising were not related to brand development in the market, encouraging marketers to employ other measures and techniques to choose geographies to target with spot advertising. Information on regional food preferences could be one component of a study on the effectiveness of spot advertising for food products.

Marketers also vary their trade and consumer promotional programs by geography and allow many sales representatives some flexibility to tailor promotions to the needs of each account. According to one survey, account-specific marketing programs have grown to about 10% of the average consumer packaged good firm's marketing budget (Spethmann, 2001). Recently, Kraft developed 150 different versions of a Sunday free-standing insert coupon advertisement so the message and offer could be tailored to local market characteristics (Kinsman, 2002). In addition, the growing interest in targeting ethnic groups, illustrated by a 44% increase in Hispanic advertising between 1998 and 2002 (Hopewell, 2003), has increased marketer interest in regional purchase patterns because ethnic groups are not uniformly distributed across the country. These examples suggest there is an emerging need for methods and analyses that find areas with similarities and note which markets have large differences.

3. METHODOLOGY AND DATA

Cluster analysis has been employed in many disciplines to address a variety of issues. It has been used to classify farms based on their proneness toward residential development (Levia & Page, 2000), to group options for water resource planners to simplify decision making (Bari, 1992), to develop a classification scheme for rose clover cultivars (Nunes & Smith, 2003), to segment business-to-business buyers based on the attributes they appreciate in Web sites (Lord & Collins, 2002), and to better understand the labor productivity and wages paid in similar industries (Galbraith, 1998). Johnson (2001) suggested cluster analysis for the identification of comparable parties for estimating internal transfer prices. Huttin (2000) grouped patients according to socioeconomic data using cluster analysis and analyzed variations in spending on prescribed medications by households. Machauer and Morgner (2001) used the technique to develop segments of bank buyers based on their attitudes and preferences and Fotopoulos and Krystallis (2003) used it to better understand the wants and needs of Zagora apple buyers in Greece. This technique has also been recently used to address many issues in the agricultural economics and agribusiness literatures (e.g., Baker & Burnham, 2002; Bessant, 2000; Dahl & Wilson, 2000; Parsons, Hanson, Musser, Freund, & Power, 2000; Pennings & Leuthold, 2000; Richards, 2000; Rudstrom, Popp, & Manning, 2002).

Although cluster analysis has become a fairly common analytical technique, some researchers do not follow procedures that, according to Monte Carlo studies, have the highest likelihood of identifying the underlying groups. For example, the Kmeans algorithm, a popular cluster analysis technique, requires analysts to specify how many clusters to identify. The actual number of groups could be larger or smaller. In addition, like many nonlinear programming algorithms, if Kmeans does not start with a good initial grouping, it may stop at a local optimum that is not the globally best cluster result. Another approach is to use hierarchical algorithms. These algorithms start with all the objects as separate clusters and combine the clusters one at a time until all the objects are in a single group. One problem with hierarchical techniques is that objects that are combined early in the process will stay together even if later groupings would be better if they were in separate clusters. Many algorithms are available and Monte Carlo studies have found that some tend to work better than others. Some marketing researchers use several clustering algorithms and check if they arrive at similar results (Shepard, 2003). Others follow a two-stage approach, using the results from one cluster analysis as the starting point for a second analysis. Milligan and Cooper (1987) summarized the findings from Monte Carlo studies and proposed a seven-step process to enhance the likelihood that a cluster analysis will identify the underlying groupings in the data. This study will follow the seven-step process and will illustrate its use for other researchers.

The first step is to select objects for the analysis. Many food marketers purchase scanner-based sales data from the ACNielsen Company to track market trends and evaluate product sales. This study will use dollar sales per capita indices for the 52-weeks ending on June 16, 2001 from the ACNielsen Company. The data are for the 52 ACNielsen Scantrack Markets illustrated in Figure 1. The market names and their abbreviations are listed in Table 1.

The next step is to select the variables. Table 2 lists the 62 food categories used in this study. These categories were developed to serve the needs of ACNielsen's clients. They represent nearly all UPC-scannable (i.e., non-random weight) food products sold in stores. Weekly scanner sales data, collected from a large sample of food, mass merchant, and

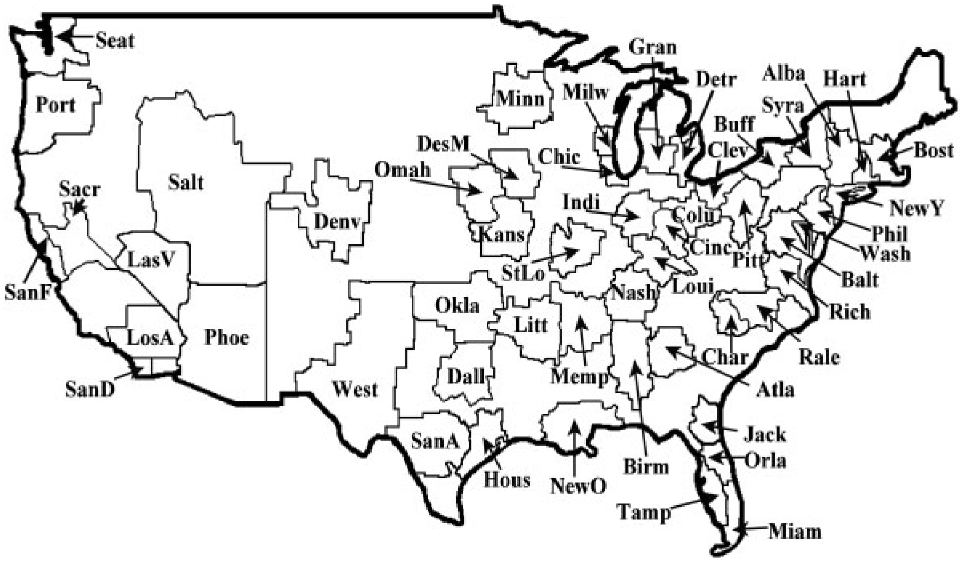

Figure 1 ACNielsen Scantrack market map.

TABLE 1. ACNielsen Scantrack Market Names and Abbreviations

Abbreviation	Market name	Abbreviation	Market name
Alba	Albany	Miam	Miami
Atla	Atlanta	Milw	Milwaukee
Balt	Baltimore	Minn	Minneapolis
Birm	Birmingham	Nash	Nashville
Bost	Boston	NewO	New Orleans/Mobile
Buff	Buffalo/Rochester	NewY	New York
Char	Charlotte	Okla	Oklahoma City/Tulsa
Chic	Chicago	Omah	Omaha
Cinc	Cincinnati	Orla	Orlando
Clev	Cleveland	Phil	Philadelphia
Colu	Columbus	Phoe	Phoenix
Dall	Dallas	Pitt	Pittsburgh
Denv	Denver	Port	Portland
DesM	Des Moines	Rale	Raleigh/Durham
Detr	Detroit	Rich	Richmond/Norfolk
Gran	Grand Rapids	Sacr	Sacramento
Hart	Hartford/New Haven	Salt	Salt Lake City/Boise
Hous	Houston	SanA	San Antonio
Indi	Indianapolis	SanD	San Diego
Jack	Jacksonville	SanF	San Francisco
Kans	Kansas City	Seat	Seattle
LasV	Las Vegas	StLo	St. Louis
Litt	Little Rock	Syra	Syracuse
LosA	Los Angeles	Tamp	Tampa
Loui	Louisville	Wash	Washington, DC
Memp	Memphis	West	West Texas/New Mexico

TABLE 2. ACNielsen Scantrack Food Categories in this Analysis

Baby Food	Frozen Breakfast Foods
Baking Mixes	Frozen Vegetables
Bottled Water	Frozen Juices and Drinks
Bread and Baked Goods	Frozen Meal Starters
Breakfast Foods	Gum
Butter and Margarine	Ice Cream
Candy	Jams/Jellies/Spreads
Canned Vegetables	Milk
Canned Fruit	Non-Carbonated Soft Drinks
Canned Seafood	Nuts
Carbonated Beverages	Packaged Meat
Cereal	Packaged Milks and Modifiers
Cheese	Pasta
Coffee	Pickles/Olives/Relishes
Condiments/Gravies/Sauces	Prepared Foods—Dry Mixes
Cookies/Ice Cream Cones	Prepared Food/Deli Salads/Dressing
Cottage Cheese/Sour Cream/Toppings	Puddings/Desserts—Dairy
Crackers	Ready-to-Serve Prepared Foods
Desserts/Gels/Syrups	Refrigerated Dough Products
Dried Fruit	Refrigerated Juices and Drinks
Dry Vegetables and Grains	Salad Dressings/Mayonnaise/Toppings
Flour	Shelf Stable Juices and Drinks
Fresh Produce	Shortening/Oil
Fresh Eggs	Snacks
Fresh Meat (with Manufacturer UPCs)	Snacks/Spreads/Dips—Dairy
Frozen Desserts/Fruits/Toppings	Soup
Frozen Novelties	Spices/Seasonings/Extracts
Frozen Prepared Foods	Sugar/Sugar Substitutes
Frozen Baked Goods	Table Syrups/Molasses
Frozen Pizza/Snacks	Tea
Frozen Unprepared Meat/Seafood	Yogurt

drug stores, is used to project total consumer purchases by market (Wal-Mart was included in the sample during this time period). Therefore, regional variations in the channels where food is purchased should not affect the sales indices.

Another application of cluster analysis is the identification of outliers. By reversing the objects and variables, it is possible to find variables that may have measurement problems. In this case, ACNielsen originally provided data on 63 food categories. A cluster analysis that grouped the categories based on their regional patterns found three categories with unique patterns: Frozen Juices and Drinks, Fresh Meat, and Ice. Regional preferences for juices and drinks vary by whether the products are refrigerated, shelf-stable, or frozen. All three juice categories are part of the analysis. The fresh meat category only included products that could be scanned. Preferences for scannable meats and store-processed or bulk (and non-scannable) meats may vary by region. ACNielsen was confident with the quality of the first two categories. However, there were some concerns about using ice category data in this analysis. Because ice bags are bulky, some store cashiers may record purchases as miscellaneous sales rather than scan the bags. If policies about scanning ice varied by store, some regional biases could show up in the data. Therefore, the ice category was dropped from the analysis.

Deciding whether to standardize the variables is the third step in the procedure. The indices show differences in food consumption by market. For example, butter and margarine purchases were highest in Syracuse (dollar sales per capita index of 152) and Albany (146) and lowest in Los Angeles (62) and West Texas/New Mexico (58). Carbonated beverage sales were highest in Des Moines (172) and Little Rock (156) and lowest in San Francisco (69) and New York (58). If a cluster analysis included variables with different scales (e.g., population and average age), the larger variable would receive more weight in the analysis. Standardization reduces the effects from variable scales. One might think standardization is not needed in this study because the variables are indexed to the U.S. average. However, Larson (1997b) illustrated that standardization can impact a cluster analysis with percentage variables. Milligan and Cooper (1988) tested different standardization procedures and concluded that dividing each variable by its range was the best. In this study, each variable is divided by its range.

The next two steps in the clustering procedure involve selecting the similarity measure and the clustering methods. Euclidean distance will be used to measure similarity. Ward's and Beta-Flexible hierarchical algorithms (with beta equal to -0.25) will be employed along with a Kmeans partitioning algorithm. The results from these high-rated algorithms will be compared in an attempt to form a consensus on the best market grouping (Bayne, Beauchamp, Begovich, & Kane, 1980; Milligan, 1980; 1989; Scheibler & Schneider, 1985). A two-stage process, using hierarchical algorithms to develop starting points or seeds for Kmeans, will also be used. By combining the techniques, starting point problems for Kmeans and path dependence problems for hierarchical algorithms are reduced. This two-stage approach has been used in several other studies (e.g., Bunn, 1993; Fournier, Antes, & Beaumier, 1992; Larson, 1997a, 1997b, 1998; Nairn & Bottomley, 2003; Wansink & Park, 2000).

The sixth step in the process involves selecting stopping rules to help guide the determination of how many clusters to use in the final solution. Milligan and Cooper (1985) and Cooper and Milligan (1988) found that the Pseudo-F and Pseudo-T^2 stopping rules work fairly well in Monte Carlo studies. The Pseudo-F statistic suggests a cluster level may be a good place to stop if it is higher than the previous and next level's statistics. The Pseudo-T^2 stopping rule suggests a level may be a good place to stop if it is lower than the previous and next level's statistics. These rules will guide the selection of the number of clusters in this study.

The final step in the clustering procedure is to interpret, test, and replicate the results. In a sense, the clustering process uses up all the degrees of freedom in the data, making statistical measures or tests of grouping quality impossible. Considerable judgment is needed to select the most reasonable set of clusters, particularly when several algorithms are used and they do not reach a consensus grouping.

4. RESULTS

The CLUSTER and FASTCLUS procedures in SAS were used to develop the candidate groupings for the data (SAS Institute, 1989). Figure 2 shows the clusters from level 13 to level 4 as a tree diagram from the Ward's algorithm and Figure 3 shows the Beta Flexible results. Both algorithms had very similar groupings at level 13. For example, in the Ward's results at level 13, Albany, Boston, Buffalo/Rochester, Chicago, Hartford/New Haven,

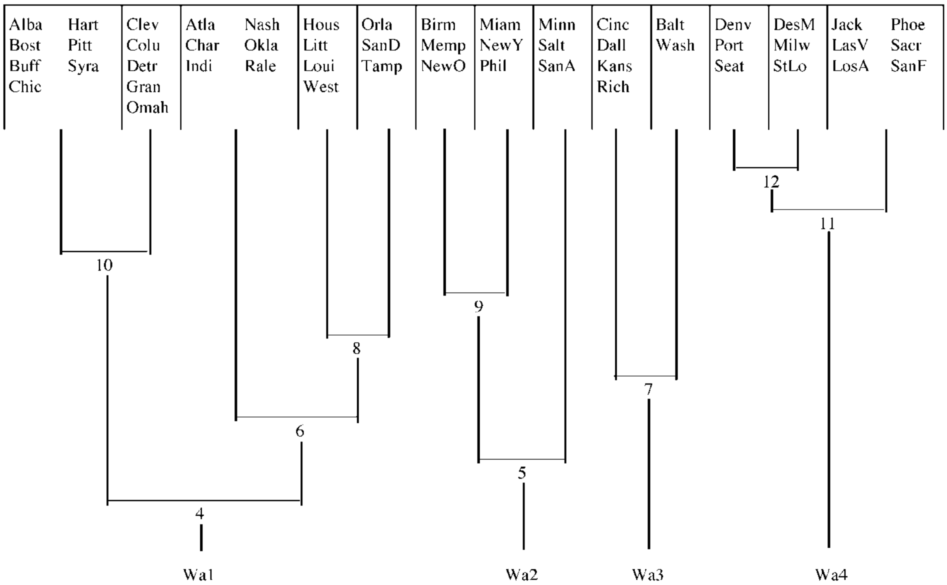

Figure 2 Clusters 13 to 4 from Ward's Algorithm, standardized dollar sales per capita index data.

Pittsburgh, and Syracuse were together in one cluster. To make 12 clusters, the group consisting of Denver, Portland, and Seattle was merged with the group consisting of Des Moines, Milwaukee, and St. Louis. At level 4, most of the Northeast and Southeast markets were merged together in a cluster (Wa1).

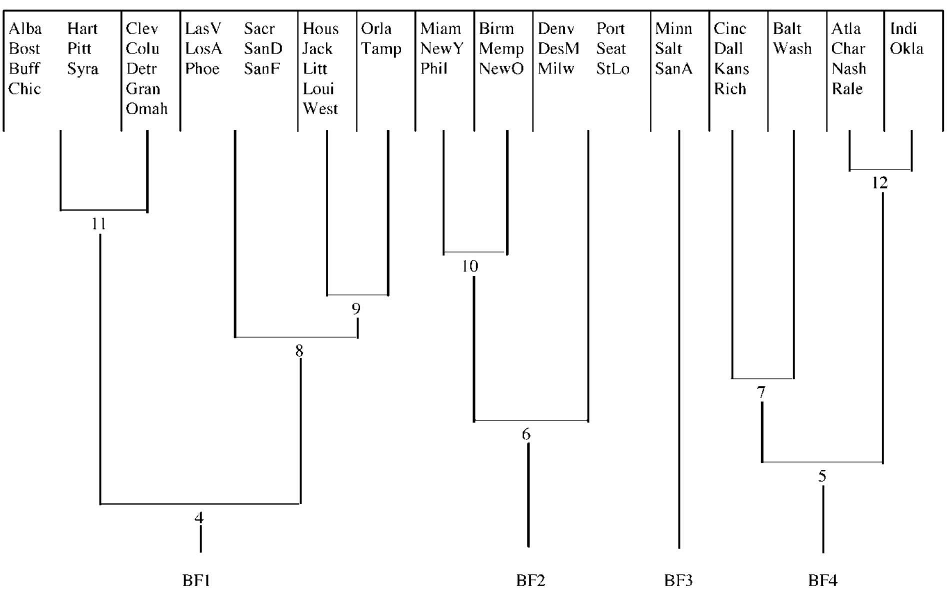

Figure 3 Clusters 13 to 4 from Beta Flexible Algorithm, standardized dollar sales per capita index data.

Table 3 shows the stopping rule results for the five different clustering processes. The Kmeans algorithm does not produce Pseudo-T^2 statistics. Pseudo-F statistics for both the Ward's and Beta Flexible algorithms suggested level 11 could be a good grouping. However, the algorithms produced different clusters at level 11. The Kmeans algorithm did not change the results from either grouping, so no consensus grouping was reached. Pseudo-T^2 statistics for the Ward's results also suggested that level 11 could be a good stopping point.

The results from the two algorithms were compared to find the most reasonable results. To create 11 groups, the Ward's algorithm combined markets from the far Northwest along with Des Moines, Milwaukee, and St. Louis with markets from the far Southwest along with Jacksonville. Although having a large Western cluster is possible, adding the other markets seems less reasonable. To create 11 groups, the Beta Flexible algorithm combined markets from the Northeast and Chicago with markets from the upper Midwest. The Beta Flexible algorithm also had San Diego linked with other markets in the far Southwest instead of with two Florida markets and Jacksonville linked with other markets in the South instead of with the far Southwest. Moving down to level 10 in the Beta Flexible results does not appear reasonable because this would combine New York, Philadelphia, and Miami with three "Deep South" markets. Since the level 11 results from the Beta Flexible algorithm were more intuitive, they were selected as the preferred clusters.

The preferred level 11 results are shown in Figures 4 and 5. Figure 4 shows that there was a far Southwest cluster and a large Northeast cluster that ranged from Chicago to Boston and included Omaha. Consumption patterns in Seattle, Portland, and Denver were similar to those in Des Moines, Milwaukee, and St. Louis. Several "Heartland" markets were a cluster, and Baltimore and Washington formed a separate cluster. Figure 5 shows some results that are more difficult to explain. The four Texas markets were split into three different clusters and the four Florida markets were split into three different clusters. Miami consumption patterns were similar to those in New York and Philadelphia. Migration patterns may have contributed to this finding. Two clusters may need additional research: the grouping of Minneapolis, Salt Lake City/Boise, and San Antonio and the grouping of Cincinnati, Dallas, Kansas City, and Richmond. Perhaps demographic

TABLE 3. Pseudo-F and Pseudo-T^2 Stopping Rule Results for Clusters 13 to 4 Based on Standardized Dollar Sales per Capita Index Data

Cluster	Ward's Pseudo F	Ward's + Kmeans Pseudo F	Beta Flexible Pseudo F	Beta + Kmeans Pseudo F	Kmeans Pseudo F	Ward's Pseudo T^2	Beta Flexible Pseudo T^2
13	16.9	16.88	16.7	16.74	12.63	3.8	2.2
12	17.5	17.49	17.4	17.36	14.54	4.2	3.8
11	**17.4**	**17.44**	**17.3**	**17.27**	15.84	**5.2**	4.6
10	17.7	17.66	17.4	17.45	**14.32**	4.6	**6.7**
9	18.2	18.18	18.1	18.13	16.92	**6.7**	3.9
8	18.8	18.77	**18.1**	**18.10**	18.00	5.0	**5.8**
7	19.8	19.81	19.1	19.14	18.12	3.6	3.6
6	21.1	21.87	20.4	20.69	20.28	**4.8**	5.7
5	23.5	24.33	20.8	21.02	23.27	4.6	7.1
4	24.1	25.25	22.1	24.76	24.45	9.7	9.3

Note. Underlined entries suggest which levels may have good clustering.

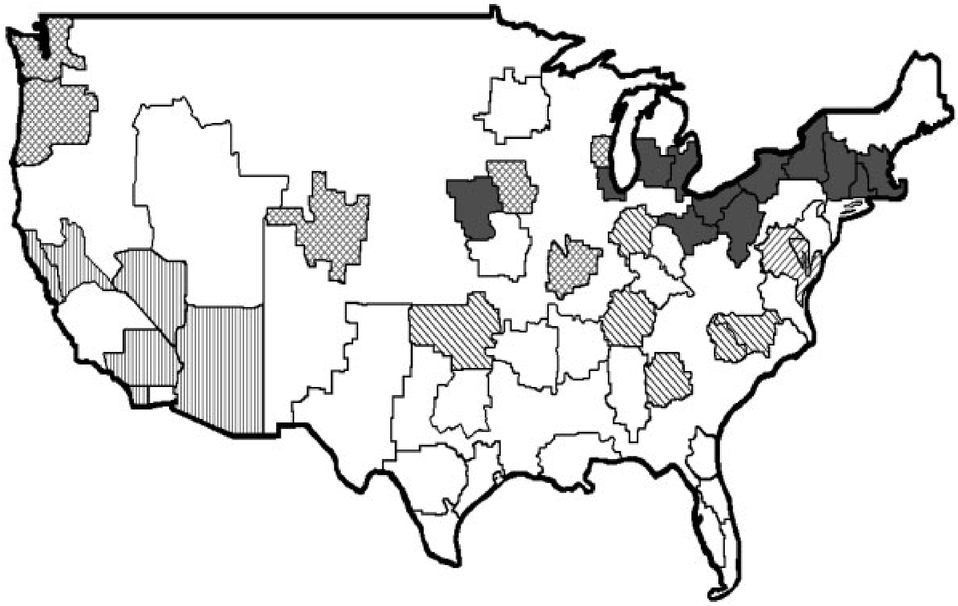

Figure 4 Preferred cluster grouping—Clusters 1 to 5 from Beta Flexible Algorithm.

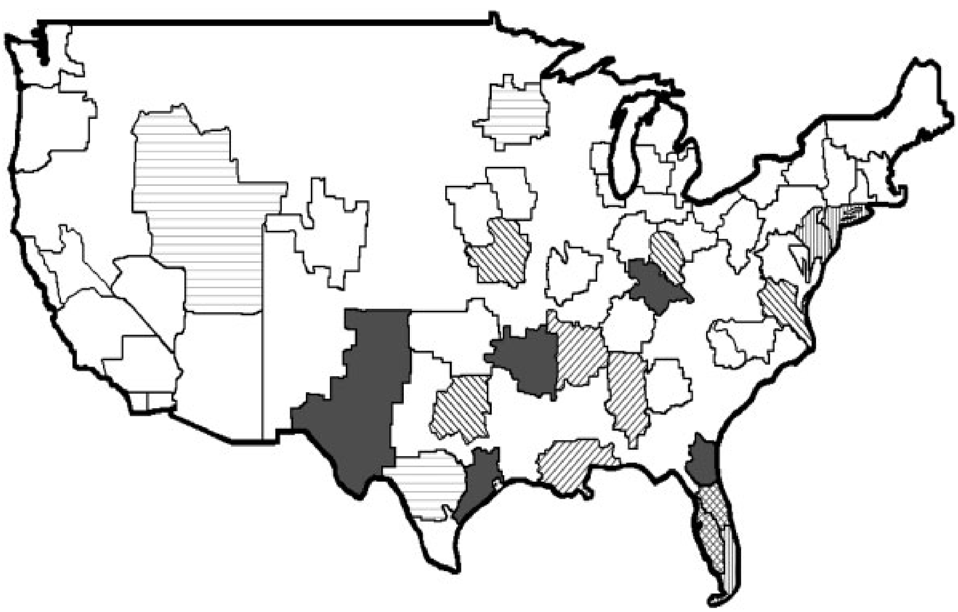

Figure 5 Preferred cluster grouping—Clusters 6 to 11 from Beta Flexible Algorithm.

profiles of those markets, store formats, or other variables can explain the similarity of the food consumption patterns.

5. STABILITY OF MARKET CLUSTERS

To explore the stability of food consumption patterns, the results from this study will be compared with those from a previous study (Larson, 1998). In that study Selling-Area Marketing, Inc., (SAMI) market data for 126 food categories was used. This data, based on warehouse withdrawals, was collected every 4 weeks. When a store ordered products, the items were withdrawn from warehouses. By working with nearly every major food warehouse in the country, SAMI tracked the withdrawals and linked the volume to where the store was located. SAMI projected food store sales by market using warehouse withdrawals, sold annual dollar sales per household (category development) indices to clients, and provided 1990 indices for the analysis.

There were some similarities between the 1990 cluster results and this study. Twenty-two of the SAMI markets can be matched up with 24 similar Scantrack markets in comparable clusters (bold entries in Table 4). There were other similarities. In the Scantrack clusters, Milwaukee was linked with Des Moines, part of the SAMI Omaha/Des Moines Market. Indianapolis and Oklahoma City/Tulsa, Grand Rapids/Kalamazoo and Pittsburgh, and Atlanta and Nashville were together in both studies. Albany, Boston, and Hartford/New Haven were also in the same cluster, just like with the SAMI results.

Larson (1998) discussed the findings of other studies that provided some external validation for the SAMI groupings. Zelinsky (1974,1987) and Reed (1976) examined a variety of measures and concluded that Florida was different from the rest of the South and that large areas often had similar tastes. Management Horizons, Inc., developed its own set of regions and concluded that the far Southwest was different from the far Northwest (Whalen, 1983). The findings from these studies also provide external support for the market groupings based on Scantrack data.

There were some differences in the market groupings. Denver moved from the far Southwest in the SAMI results to the far Northwest in the Scantrack clusters. Salt Lake City was no longer linked with a cluster of Western markets. Texas's four markets were in the same SAMI cluster, but were divided into three clusters in the Scantrack results. Miami was still linked with New York and Philadelphia, but the rest of Florida was divided into two other clusters. Jacksonville was grouped with Houston, West Texas/New Mexico, Little Rock, and Louisville while food consumption patterns in Orlando and Tampa were unique enough for them to remain a separate cluster.

Other researchers have examined changes in Florida. Lamme and Meindl (2002) found significant shifts over time in the areas with high concentrations of Northeast-born and Midwest-born residents. This could explain at least some of the cluster results from Florida.

Differences in the geographic coverage of the SAMI markets and Scantrack markets should be considered. Of the 54 SAMI markets, Scantrack did not cover eleven of them: Spokane, Wichita, Green Bay, Quad Cities, Peoria-Springfield, Shreveport-Jackson, Charleston-Savannah, Greenville-Spartanburg-Asheville, Charleston-Huntington, Scranton-Wilkes Barre, PA, and Portland, ME. Scantrack covered Las Vegas and split Sacramento from San Francisco, San Diego from Los Angeles, Washington, DC from Baltimore, Little Rock from Memphis, Columbus from Cincinnati, Des Moines from Omaha, and Orlando and Tampa from Jacksonville. Since many of these divided markets were clustered together,

TABLE 4. Comparison of Cluster Results (Boldfaced Entries Show a Market Match)

Cluster	Based on SAMI warehouse withdrawals (Larson, 1998)		Based on ACNielsen Scantrack	
1	**Portland, OR** Salt Lake City/Boise	**Seattle/Tacoma** Spokane/Yakima, WA	Denv Milw **Seat**	DesM **Port** StLo
2	Denver **Phoenix/Tucson**	**Los Angeles/San Diego** **San Francisco**	LasV **Phoe** **SanD**	**LosA** **Sacra** **SanF**
3	**Buffalo/Rochester** **Chicago** **Detroit**	Baltimore/Washington **Cleveland** **Syracuse**	Alba **Buff** **Chic** **Detr** Hart Pitt	Bost **Clev** Colu Gran Omah **Syra**
4	Green Bay Omaha/Des Moines Quad Cities	Milwaukee Peoria/Springfield, IL	Orla Tamp	
5	**Cincinnati/Dayton/Columbus** Indianapolis Louisville/Lexington, KY St. Louis	Charleston/Huntington **Kansas City** Oklahoma City/Tulsa Wichita	**Cinc** **Kans**	Dall Rich
6	Atlanta Shreveport/Jackson	Nashville/Knoxville, TN	Balt	Wash
7	Birmingham/Montgomery/Huntsville Charleston/Savannah Norfolk/Richmond	**Charlotte** Greenville/Spartanburg/Asheville **Raleigh/Greensboro/Winston Salem**	Atla Indi Okla	**Char** Nash **Rale**
8	Dallas/Fort Worth **Houston**	**El Paso/Albuquerque/Lubbock** San Antonio/Corpus Christi	**Hous** Litt **West**	Jack Loui
9	**Memphis/Little Rock**	**New Orleans**	Birm **Memp** **NewO**	
10	Grand Rapids/Kalamazoo Pittsburgh	**Minneapolis/St. Paul** Portland, ME	**Minn** SanA	Salt
11	Albany/Schenectady/Troy Hartford/New Haven/Springfield, CT **Miami** **Philadelphia**	Boston/Providence Jacksonville/Orlando/Tampa **New York City** Scranton/Wilkes Barre, PA	**Miam** **NewY** **Phil**	

they helped confirm the existence of regional food consumption patterns. In several cases, the market dimensions also differed between the two data sets (e.g., the Salt Lake City/Boise market was larger for Scantrack). The effect from these differences, greater coverage in some areas and less detail in other areas, should have little impact on the overall geographic patterns of food consumption.

Several differences in the data sets might have contributed to the variations in the results. SAMI used warehouse withdrawals to track sales in grocery stores. ACNielsen used scanner data to track sales in the food, mass merchant, and drug channels. Because supercenters were not common in the food industry in 1990, the narrower coverage by SAMI probably had little impact on the results. Given the growth of food sales through channels other than supermarkets, the broader coverage from Scantrack is needed to track regional sales patterns today. The SAMI cluster analysis used annual dollar-sales-per-household indices while the Scantrack analysis used annual dollar-sales-per-capita indices. This factor might affect markets with particularly large and small household sizes. Perhaps the high average household size in Salt Lake City/Boise market contributed to the shift from the far Northwest to a different cluster with the Scantrack groupings. Another factor is the level of detail in the study. The SAMI analysis used data from 126 food categories and the Scantrack analysis used data from 62 larger food categories. Categories with significant direct store delivery were dropped from the SAMI analysis because their volume did not pass through warehouses and was not tracked. Sales of these products were tracked by scanners, so their volume is counted by Scantrack. Although the Scantrack data had half as many categories, it covered a larger portion of total food sales. Because the Scantrack categories combined several SAMI categories, some regional preferences may not be represented. Similarities between the SAMI and Scantrack market groupings suggest that the information loss may not be significant. The effects of this aggregation could be tested by repeating this analysis with category segments. The final factor is the change in the time frame. Since the other factors do not appear to offer explanations for many variations in the results, some regional food preference shifts probably occurred during the 10-year period. The changes in Texas and Florida suggest that the population fragmentation Weiss referred to may be reflected in changes in the regional patterns of food consumption.

6. SUMMARY AND IMPLICATIONS

Regional variations in food consumption continue to be important. The new market groupings revealed many areas that have similar consumption patterns. Food consumption was fairly similar in the far Northwest cluster, in the far Southwest cluster, and in a group of markets that ranged from Chicago to Boston. These 11 clusters may help those who work with scanner-based data control for differences in food preferences across the US. The cluster results may be used to help forecast product sales after a regional product introduction because a product's average sales rate across a cluster may be similar to its sales rate in part of the cluster. The technique can be used to help develop homogeneous sales territories and to identify areas that may benefit from extra marketing support. Sales deviations within a cluster could suggest opportunities. Products and promotions can be tailored to the preferences of the people who reside in each cluster. The new groupings can be employed by managers and policy makers to benchmark performance and test the effectiveness of different marketing and educational initiatives. Marketers that use focus groups may want to host groups in each cluster to learn more about preference differences across the country. The clusters can help balance food preferences in test and control cells to improve the quality of research experiments.

The new groupings had many similarities with the cluster analysis based on 1990 data, further strengthened the case against using regions from 1910 in cross-section analyses. Food consumption patterns in the Western and Southern regions are not homogeneous. Instead of grouping the Pacific markets together, they should continue to be separated into the far Northwest and the far Southwest. The South appears to have very different food consumption patterns as one moves from East to West. Florida is quite different from the rest of the South with Miami's consumption patterns being similar to those in New York City and Philadelphia. Some changes occurred during the 10-year period including the fragmentation of some market areas (e.g., Texas and Florida) and some shifts in the market groupings (e.g., Denver and Salt Lake City/Boise). Therefore, researchers, managers, and policy makers should regularly update their regional groupings to be sure they reflect current geographic purchase patterns.

ACKNOWLEDGMENTS

The author wishes to thank Jingcai Chen for her research assistance and the ACNielsen Company for graciously providing the data for this analysis.

REFERENCES

Baker, G.A., & Burnham, T.A. (2002). The market for genetically modified foods: Consumer characteristics and policy implications. International Food and Agribusiness Management Review, 4(4), 351–360.
Bari, M.F. (1992). Cluster analysis as an aid in water resources planning and decision making. Irrigation Engineering and Rural Planning, 22, 29–42.
Bayne, C.K., Beauchamp, J.J., Begovich, C.L., & Kane, V.E. (1980). Monte Carlo comparisons of selected clustering procedures. Pattern Recognition, 12(2), 51–62.
Bessant, K.C. (2000). Part-time farming situation among Manitoba farm operators: A typological approach. Canadian Journal of Agricultural Economics, 48(3), 259–277.
Blank, C. (1998). McDonald's McMenu is bolstering its regional offerings. Supermarket News, 48(40), 25, 34.
Bunn, M.D. (1993). Taxonomy of buying decision approaches. Journal of Marketing, 57(1), 38–56.
Cardona, M.M. (1998). Study finds half of spot TV buys are wasted money. Advertising Age, 69(13), 22.
Cooper, M.C., & Milligan, G.W. (1988). The effect of measurement error on determining the number of clusters in cluster analysis. In W. Gaul & M. Schrader (Eds.), Data, expert knowledge and decisions: An interdisciplinary approach with emphasis on marketing applications (pp. 319–328). Berlin/New York: Springer-Verlag.
Dahl, B.L., & Wilson, W.W. (2000). Grades/classes of hard wheats exported from North America: Analysis of growth and market segments. Review of Agricultural Economics, 22(1), 172–191.
Fotopoulos, C., & Krystallis, A. (2003). Quality labels as a marketing advantage: The case of the 'PDO Zagora' apples in the Greek market. European Journal of Marketing, 37(10), 1350–1374.
Fournier, S., Antes, D., & Beaumier, G. (1992). Nine consumption lifestyles. In J.F. Sherry, Jr. & B. Sternthal (Eds.), Advances in Consumer Research, 19, 329–337.
Galbraith, J.K. (1998). Created unequal: The crisis in American pay. New York: The Free Press.
Gerrie, D. (1987, August 27). Area of choice. Marketing, pp. 31–32.
Grewal, D., Levy, M., Mehrotra, A., & Sharma, A. (1999). Planning merchandising decisions to account for regional and product assortment differences. Journal of Retailing, 73(3), 405–424.
Grigg, D. (1999). Food consumption in the Mediterranean region. Journal of Economic and Social Geography, 90(4), 391–409.
Harris, E.W., & Nowverl, A. (2000). What's happening to soul food? Regional and income differences in the African American diet. Ecology of Food and Nutrition, 38(6), 587–603.
Hiam, A., & Schewe, C.D. (1992). The portable MBA in marketing. New York: Wiley.
Hopewell, N. (2003). 2003 marketing fact book. Marketing News, 37(14), 15–24.
Huttin, C. (2000). A cluster analysis on income elasticity variations and U.S. pharmaceutical expenditures. Applied Economics, 32(10), 1241–1247.
Johnson, R.E. (2001). The role of cluster analysis in assessing comparability under the U.S. transfer pricing regulations. Business Economics, 36(2), 30–38.
Kinsman, M. (2002). Retailgating: Today's promotions are one part national, 150 parts local. Promo, 15(4), 24–25.
Lamme, A.J., & Meindl, C.F. (2002). A vibrant cultural boundary in Florida. Southeastern Geographer, 42(2), 274–295.

Larson, R.B. (1997a). Food consumption and seasonality. Journal of Food Distribution Research, 28(2), 36–44.
Larson, R.B. (1997b). Clustering and balancing markets with demographic data. Journal of Food Products Marketing, 4(4), 3–24.
Larson, R.B. (1998). Regionality of food consumption. Agribusiness, 14, 213–226.
Leeds, D.B. (1994). Point-of-purchase. In T.B. Block & W.A. Robinson (Eds.), The Dartnell sales promotion handbook (pp. 166–188). Chicago, IL: The Dartnell Corporation.
Levia, D.F., Jr., & Page, D.R. (2000). The use of cluster analysis in distinguishing farmland prone to residential development: A case study of Sterling, Massachusetts. Environmental Management, 25(2), 541–548.
Linneman, R.E., & Stanton, J.L. (1991). Making niche marketing work. New York: McGraw-Hill.
Lord, K.R., & Collins, A.F. (2002). Suppler web-page design and organizational buyer preferences. Journal of Business and Industrial Marketing, 17(2/3), 139–150.
Machauer, A., & Morgner, S. (2001). Segmentation of bank customers by expected benefits and attitudes. International Journal of Bank Marketing, 19(1), 6–17.
McKenna, S. (1992). The complete guide to regional marketing. Homewood, IL: Business One Irwin.
Milligan, G.W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. Psychometrika, 45(3), 325–342.
Milligan, G.W. (1989). A study of the beta-flexible clustering method. Multivariate Behavioral Research, 24(2), 163–176.
Milligan, G.W., & Cooper, M.C. (1985). An examination of procedures for determining the number of clusters in a data set. Psychometrika, 50(2), 159–179.
Milligan, G.W., & Cooper, M.C. (1987). Methodology review: Clustering methods. Applied Psychological Measurement, 11(4), 329–354.
Milligan, G.W., & Cooper, M.C. (1988). A study of standardization of variables in cluster analysis. Journal of Classification, 5(2), 181–204.
Mood, F. (1951). The origin, evolution, and application of the sectional concept, 1750–1900. In M. Jensen (Ed.), Regionalism in America (pp. 5–98). Madison, WI: The University of Wisconsin Press.
Mulhern, F.J., Williams, J.D., & Leone, R.P. (1998). Variability of brand price elasticities across retail stores: Ethnic, income, and brand determinants. Journal of Retailing, 74(3), 427–446.
Nairn, A., & Bottomley, P. (2003). Something approaching science? Cluster analysis procedures in the CRM era. International Journal of Market Research, 25(2), 241–263.
Nunes, M.E.S., & Smith, G.R. (2003). Characterization of rose clover germplasm for flowering traits. Crop Science, 43(4), 1523–1527.
Parsons, R.L., Hanson, G.D., Musser, W.N., Freund, R., & Power, L. (2000). A financial training program for USDA/FSA borrowers: Evolution and impacts. Agricultural and Resource Economics Review, 29(2), 240–250.
Pennings, J.M.E., & Leuthold, R.M. (2000). The role of farmers' behavioral attitudes and heterogeneity in futures contracts usage. American Journal of Agricultural Economics, 82(4), 908–919.
Reed, J.S. (1976). The heart of Dixie: An essay in folk geography. Social Forces, 54(4), 925–939.
Richards, T.J. (2000). A discrete/continuous model of fruit promotion, advertising, and response segmentation. Agribusiness, 16, 179–196.
Rudstrom, M., Popp, M., & Manning, P. (2002). Data aggregation issues for crop yield risk analysis. Canadian Journal of Agricultural Economics, 50(2), 185–200.
SAS Institute. (1989). SAS/STAT user's guide, Version 6.1, 4th ed. Cary, NC: SAS Institute, Inc.
Scheibler, D., & Schneider, W. (1985). Monte Carlo tests of the accuracy of cluster analysis algorithms: A comparison of hierarchical and nonhierarchical methods. Multivariate Behavioral Research, 20(3), 283–304.
Shepard, D. (2003). How to tackle a cluster analysis. Direct, 15(7), 15, 41.
Spethmann, B. (2001). Trade promotion: Going for broke. Promo, 14(9), 27–31.
Urbanski, A. (2003). The merchandising emergency kit. Progressive Grocer, 82(5), 21–34.
Wansink, B., & Park, S.B. (2000). Methods and measure that profile heavy users. Journal of Advertising Research, 40(4), 61–72.
Weinstein, A. (1987). Market Segmentation: Using demographics, psychographics and other segmentation techniques to uncover and exploit new markets. Chicago, IL: Probus Publishing Co.

Weiss, M.J. (1988). The clustering of America. New York: Harper & Row.
Weiss, M.J. (1994). Latitudes and attitudes: An atlas of American tastes, trends, politics, and passions. Boston, MA: Little, Brown & Co.
Weiss, M.J. (2000). The clustered world: How we live, what we buy, and what it all means about who we are. Boston, MA: Little, Brown & Co.
Whalen, B. (1983). Marketing research firm segments U.S. into 15 geodemographic regions. Marketing News, 17(10), Section 2, 8.
Yeager, D. (1987). Varying buying patterns spur regional marketing. Supermarket News, 37(37), 16–18.
Zelinsky, W. (1974). Selfward bound? Personal preference patterns and the changing map of American society. Economic Geography, 20(2), 144–179.
Zelinsky, W. (1987). You are where you eat. American Demographics, 9(7), 30–33, 56–61.

Ronald B. Larson *is an Associate Professor in the Department of Marketing in the Haworth College of Business at Western Michigan University in Kalamazoo, MI. He received his PhD in agricultural economics from Purdue University in 1993. His current research interests include food marketing, sales promotion, and marketing strategy.*