
Consumers' Retail Source of Food: A Cluster Analysis

Andrea Carlson, PhD
U.S. Department of Agriculture
Center for Nutrition Policy and Promotion

Jean Kinsey, PhD
University of Minnesota

Carmel Nadav, PhD
Venturi Technology Partners

The popular impression that only half of our food comes from retail grocery stores is based on food expenditure data. However, the U.S. Department of Agriculture's Continuing Survey of Food Intakes by Individuals, 1994 shows that 72 percent of the amount of food (measured in grams) consumed by Americans comes from grocery stores. Using cluster analysis, we grouped consumers based on where they obtained their food and found that half were "Home Cookers"—purchasing 93 percent of their food from grocery stores. By comparison, the "High Service" consumers, which represented 10 percent of the sample, purchased 43 percent of their food from restaurants. This research quantifies the different shopping behaviors exhibited by groups of people in the United States and discusses some of the demographic differences among the clusters. The results are of interest to consumers, nutrition counselors, food retailers, and policymakers who deal with retail food, low-income diets, or food safety.

In economic analysis of consumer behavior, substituting expenditure for quantity is a common practice. For example, expenditure is often substituted for quantity when estimating the percentage change in the amount consumed when income changes by 1 percent (Engel function). This substitution is often used because expenditure data rather than quantity are more frequently available. And from a business perspective, expenditures are more closely related to sales—the indicator (or metric) most used by businesses to measure demand for their products. Tracking consumers' food consumption behavior with expenditure data is no exception: the percentage of income spent on food is a common measure of economic well-being both for individual households and for nations.

The percentage of personal disposable income spent on food by American consumers decreased from 25 to 11

percent between 1960 and 1997 (Putnam & Allshouse, 1996). The composition of those expenditures changed noticeably, with a decreasing proportion of each food dollar being spent on food from a retail food store called "food at home." Food-away-from-home expenditures, according to the food service and restaurant sector, grew from 26 to 45 percent of each food dollar between 1960 and 1994; by the end of 1995, the amount reached 47 percent (Putnam & Allshouse, 1996). In recent years, expenditures on food away from home have approached 50 percent (Putnam & Allshouse, 1996).

The rapid rise in food-away-from-home expenditures is reflected in another metric: the high growth in sales at commercial food service establishments relative to the growth in sales in retail food stores. Between 1987 and 1999, inflation-adjusted sales in eating and drinking establishments grew an average of 2.2 percent; similar sales in

retail food stores, however, decreased an average 0.1 percent (Food Institute, 1997).

Focusing on the proportion of the food dollar that is spent in places other than a grocery store leads to the common belief that Americans eat almost half of their food away from home. The *amounts of food* consumers eat at home or away from home, however, varies considerably from the expenditure proportions reported in the literature. Expenditures in food service establishments reflect higher costs of labor (about 30 percent of the menu price), entertainment, and service.

In contrast, we reported in 1998 that when food consumption is measured in grams, the amount of food purchased from retail stores is 72 percent of all food consumed (Carlson, Kinsey, & Nadav, 1998). Another 14 percent of food (in grams) was consumed from carryout establishments (e.g., fast-food, pizza, and sandwich shops) and other restaurants combined. The remaining 14 percent came from other sources—other people and gifts, cafeterias, vending machines, coffee or food on a common tray in an office, bars and taverns, home gardens or hunting and fishing, and public programs. When food consumption is measured by expenditure, the amount of food (g) consumed away from home is 47 percent, almost twice as much as that consumed from restaurants, carryouts, and other establishments.

Our earlier research also found that where people purchase their food did not necessarily predict where they consumed their food. For example, 10 percent of food purchased in stores was not consumed at home, while 24 percent of carryout food was consumed at home (Carlson, Kinsey, & Nadav, 1998). Rising household incomes and fewer hours for household labor foretell a rising value of time and, in turn,

predict that consumers will purchase more labor services in their pursuit of food (Kinsey, 1983). Even within a grocery store, sales of ready-to-eat foods—including those that must be heated—are rising while sales for basic ingredients are falling.

Studies in the 1970's and 1980's found that higher incomes led consumers to spend more money on meals eaten out but did not necessarily lead consumers to eat more meals away from home (Prochaska & Shrimper, 1973). A similar conclusion from other research suggests that households with wives who work part-time increased their expenditures on food away from home more so than did households where wives worked full-time even though both households had the same income (Kinsey, 1983).

As women's time in the labor market expands from zero to part-time, increases in income may expand the opportunity to eat out. But as employment becomes full-time, less time is available to eat out or cook at home. Thus, continued increases in income are not further associated with increased expenditures on food away from home. In fact, increases in income may even decrease expenditures on food away from home as consumers substitute fast-foods or take-out foods for more leisurely dining away from home (Kinsey, 1983). These findings suggest that the traditional labels of "food at home" and "food away from home," as well as the use of expenditure as the metric for quantity, do not provide a complete understanding of today's consumer.

The research reported here investigates the amount of food (g) that consumers reported eating in 1994 from various retail sources and examines the common characteristics of consumers whose retail sources of food vary from the average. We used data from the USDA

Continuing Survey of Food Intakes by Individuals, 1994 (CSFII) (USDA, 1994). We examined two questions: (1) What are the unique characteristics of people who shop for food in different types of establishments? (2) How can this information be used by managers of these establishments and public policymakers? To answer these questions, we used cluster analysis to group consumers by the retail source of their food and to describe their common shopping and eating habits.

Data and Methods

The CSFII is conducted by the Agricultural Research Service (USDA, 1994).¹ We used data from 1994 because they were the most recent data available when this study began. The CSFII data provide a better picture of overall consumption behavior than do data collected at the market level where sales are the unit of measure. The CSFII reports all food eaten by 5,589 individuals in 2,540 households in the United States. Each individual reports food intake for 2 nonconsecutive days, yielding more than 150,000 observations on individual food items. For every food item, the respondent also lists the source from which the food was obtained and how much was eaten. The sources of food used in this analysis include stores, carryout restaurants, restaurants, other people, bars and taverns, cafeterias, common coffee pots or trays, vending machines, mail order, public programs, and home-grown or caught food (see box). The response rate for the CSFII is 80 percent for the first day and 76 percent for the second day. Sample weights are used in this analysis, and the results are generalizable to the population.

¹These data are available from the U.S. Department of Commerce, Technology Administration, National Technical Service, 5285 Port Royal Road, Springfield, VA 22161, (703) 487-4650, <http://www.ntis.gov>.

Analysis

The first step in our analysis was to calculate the percentage of food, measured in grams, each person consumed from each source. Cluster analysis is used to place the adult sample² into groups based on where they obtained their food. In this case, the cluster variables are the percentage of food (g) adults consumed that come from various sources. For example, if one person's diet contains 80 percent of food from stores, 5 percent from carryout restaurants, 10 percent from restaurants, and the remaining 5 percent from cafeterias, cluster analysis uses these percentages to place that person into a group with others who have similar consumption patterns.

This analysis uses the "k-means" method of clustering that is used by SAS FASTCLUS. This method is one of the better techniques available for clustering large data sets where the goal is to divide respondents into manageable and meaningful groups to describe behavior (Hartigan, 1985; SAS Institute, 1989). K-means selects the centers of the initial clusters from the first observations in the data set and then assigns the other observations to the *nearest* cluster. When an observation is added to the cluster, k-means recalculates the mean of the cluster variables, and this mean becomes the new cluster-center. If this recalculated cluster-center changes another cluster that is closest to an observation already in the cluster, then k-means moves that observation to the closest cluster and recalculates the center of its new cluster. The process continues until the number of changes is very small.

²Because children's eating behaviors are somewhat dictated by their parents, children are not included in the cluster analysis.

Categories of Food Sources

Store: supermarket, grocery store, warehouse, convenience store, drug store, gas station, bakery, deli, seafood shop, ethnic food store, health food store, commissary, produce stand, and farmers' market.

Carryout: traditional hamburger, chicken, and carryout pizza restaurants; and other restaurants where customers order, pick up, and pay for food at a counter.

Restaurants: any other establishment where the food is served at the table by restaurant staff.

Other People: food received as a gift or while a guest in someone's home.

Bars and Taverns: a location the respondent classified as a bar or tavern rather than as a restaurant, carryout restaurant, or cafeteria.

School and Non-School Cafeterias: Most non-school cafeterias are based in offices. For most of the analysis, school and non-school cafeterias are separated but are often put together in summary tables.

Common Coffee Pot or Food Tray: office coffee pots, food platters at a reception or in an office, and potluck dinners.

Vending Machines: food purchased from vending machines located within stores, restaurants, cafeterias, offices, or other locations.

Mail Order: food received from a mail order catalog or club that sends food out regularly, such as a fruit-of-the-month club.

Public Programs: a combination of several CSFII categories including child and adult care centers, day care centers in private homes, soup kitchens, shelters, food pantries, Meals on Wheels, other community food programs, and residential care facilities.

Home-Grown or Caught: food that is grown or gathered by the respondent or someone the respondent knows; meat and fish procured by hunting or fishing.

Representing 75 percent of the adult sample, six of the nine clusters get more food from stores than any other source . . .

The resulting clusters are based on 2 nonconsecutive days of dietary recall. Thus, if an individual had been sampled on a different day, he or she might have ended up in a different cluster. However, because this data set is designed to be nationally representative, similar clusters would form on any day, except major national holidays.

To reduce the bias towards observations that appear at the beginning of the data set, we used a technique recommended by SAS (SAS Institute, 1989). In the first pass, the SAS procedure forms 50 clusters and saves the cluster centers in a file. Over half of these clusters have fewer than five observations, and the centers are ignored. The remaining 24 centers form the “seeds” in the next iteration to form 24 new clusters. In the third iteration, the center of the smallest cluster is removed, and the SAS procedure forms 23 new clusters from all observations. This process continues until there are five clusters. The process is described in more detail elsewhere (Carlson, Kinsey, & Nadav, 1998; MacQueen, 1967).

The second step compared each cluster with the rest of the sample to address the two research questions. Because most of the data were categorical, this study used three nonparametric tests: the chi-squared, the Kolmogorov-Smirnov test, and the Kruskal-Wallis test (described in detail elsewhere) (Siegel, 1956). These tests measure differences in distributions of variables among different subgroups. The chi-squared test was used as an initial test for differences. Differences between the observed versus expected distributions were confirmed by the other two tests. The Kolmogorov-Smirnov test was used to measure differences between two clusters in the distribution of categorical variables that cannot be ranked (e.g., race) and the Kruskal-Wallis test for differences in categories that can be

ranked (e.g., age, income, and education). For these tests, we divided the continuous variables into categories. For example, the categories for age were 19-30, 31-40, 40-50, 50-60, 60-64, and 65+; for education, less than high school, high school degree or GED, some college, 4-year degree, and professional or graduate study.

Results and Discussion

Nineteen clusters formed around the various sources of food. Several sources, such as carryout, had more than one cluster form around it. This paper will discuss only nine of these clusters, some with names based on the unique characteristics of the cluster: Working Family, Young Professional, Manager, and City Office. In other cases, the names are based on where the people in the cluster shopped: Home Cookers, Carryout, High Service, Office, and Students and Faculty.

Sociodemographic Characteristics of the Sample

Almost half (49 percent) of the adult sample was in the Home Cookers cluster (table 1), followed by those in the Working Family cluster (11 percent), and High Service cluster (10 percent). Fewer adults were in the other clusters: Carryout, Office, Manager, Young Professional, City Office, and Students and Faculty (from 3 percent to a low of 0.6 percent).

Age, Race, and Gender. With an average age of 51, people in the Home Cookers cluster were significantly older than the rest of the adult sample (tables 1 and 2). However, the standard deviation for their age was the largest (17.9, not shown), indicating a bigger spread in age than was the case for the other clusters. Three clusters—Students and Faculty, Carryout, and Young Professional—had the youngest members (mean age of 37, 36, and 31,

Table 1. Statistically significant demographic characteristics of select clusters of consumers based on where they purchased food

Cluster	Percent of adults ¹	Age, race, and gender	Income and education	Occupation and employment	Region, urban, and household size
Home Cookers	49	Older**	Lower income** Less college**	Fewer professional/ technical, and manager/proprietor** More not employed**	
Working Family	11	Younger**	More "some college**"	More full- and part-time**	Larger households**
Carryout	3	Younger than Working Family** Fewer White*		More full- and part-time*	More Northeast**
Young Professional	0.7	Younger than Carryout**	Higher income** More college and graduate study**	More full-time**	
High Service	10	More White** More men*	Higher income** More college**	More professional/ technical, and manager/proprietor* More full-time**	
Office	2.5			More full-time**	
Manager	2.0		Higher income** More college/university**	More professional/ technical, and manager/proprietor* More full-time**	More central city*
City Office	1.0			More full-time**	More central city*
Students and Faculty	0.6	More Asian/Pacific and "other"*** Fewer females*	More college and graduate**	More full- and part-time*	More Northeast**

¹Percents do not add to 100, because all clusters are not shown in the table.

* p<.05; ** p<.01: The distribution between the cluster and the rest of the adult sample is significantly different based on the Kruskal-Wallis test.

respectively). Whereas significantly more Whites were in the High Service cluster, fewer Whites were in the Carryout cluster, and more Asian/Pacific Islanders and others were in the Students and Faculty cluster.

The High Service cluster had significantly fewer women (46 percent),³

³Differences are in the distributions between the cluster and the total adult sample. The p-values do not indicate how these distributions differ, only that they are different.

compared with the remainder of the adult sample. The Young Professional cluster also had relatively few women (35 percent), but the difference from the adult sample was not significant. The Young Professional cluster, however, represented only 0.7 percent of the total sample; thus, the small size of this cluster may have contributed to the lack of statistical significance.

Income, Education, and Employment.

Mean income among the clusters ranged from \$32,554 to \$49,072. Compared with the rest of the sample, the Home Cookers cluster had a significantly lower income; three clusters had a higher income: High Service (\$42,767), Young Professional (\$48,507), and Manager (\$49,072). Although people in the Working Family and Carryout clusters earned a household income close to the Home Cookers' income (\$36,466 and

Table 2. Basic sociodemographic characteristics of select clusters of consumers based on where they purchased food

Cluster	Adult sample	Women	Center city resident	High school degree or more	Employed	Household size	Age	Income
	Percent					Mean		
Entire Adult Sample	100	49.8	33.3	76.6	57.5	2.9	48.3	\$35,298
Home Cookers	49.0	51.4	33.4	71.0	46.5	2.9	51.4	32,554
Working Family	10.0	47.8	31.4	82.7	65.7	3.2	41.8	36,466
Carryout	11.0	45.2	40.9	82.6	78.3	3.2	36.0	34,555
Young Professional	3.0	34.8	30.4	91.3	95.7	3.4	30.8	48,507
High Service	0.7	45.5	35.1	85.6	62.1	2.8	48.3	42,767
Office	2.4	55.7	39.2	76.0	73.4	3.0	49.0	39,824
Manager ¹	2.0	45.6	45.6	82.5	77.2	2.7	46.8	49,072
City Office	0.7	52.2	52.2	82.6	91.3	2.8	41.5	35,963
Students and Faculty	1.0	68.8	21.9	90.6	87.5	3.2	36.8	44,361

¹Includes a high concentration of professionals, technical workers, managers, and proprietors.

\$34,555, respectively), the distribution of incomes in the Working Family and Carryout clusters did not differ significantly from the rest of the sample.

Educational patterns tended to follow income patterns. Whereas the Home Cookers cluster had a significantly lower educational level, compared with the total sample, several other clusters had higher levels of education: Young Professional, Students and Faculty, High Service, Manager, and Working Family. The Young Professional and Students and Faculty cluster each had more people with 4-year college degrees and graduate or professional degrees. About 83 percent each of the members of the Working Family, Carryout, City Office, and Manager clusters graduated from high school or received more education. Of these, only the Manager cluster, with more members receiving college and university degrees, had a distribution that was significantly different from the sample. Although not significantly different from the rest of the sample,

76 percent of those in the Office cluster had a high school degree or more.

Occupation and Employment. The Home Cookers cluster, compared with the High Service and Manager clusters, had significantly fewer people in professional/technical occupations or who worked as managers/proprietors. Compared with other clusters, the Home Cookers cluster was significantly more likely to have unemployed members—and a concentration of unemployed people (including retirees). Whereas only 47 percent of the people in the Home Cookers cluster were employed, most of the people in the Young Professional cluster were employed (96 percent). A little more than three-fourths of those in the Manager cluster were employed (77 percent).

Region, Urbanization, and Household Size. Two clusters, Carryout as well as Students and Faculty, were more likely than other clusters to reside in the Northeast. Two clusters, Manager and City Office, had a higher percentage of people living in center cities, 46 and

52 percent, respectively. Household⁴ size among all the clusters ranged from an average of 2.7 to 3.4. Only the distribution for the Working Family cluster differed significantly from the rest of the sample. The Carryout and Young Professional also appeared to have larger households (3.2 and 3.4, respectively), but the distributions were similar to the remainder of the adult sample.

Food Sources

Representing 75 percent of the adult sample, six of the nine clusters get more food from stores than any other source: Home Cookers (93 percent), Office (73 percent), Working Family (70 percent), Students and Faculty (54 percent), Manager (53 percent), and High Service (47 percent) (table 3).

When using grams of food rather than expenditure as a measure of consumer buying behavior, we found that stores

⁴This analysis did not include children, but we did examine the number of children present in the households of the adult respondents.

Table 3. The percentage share of food source for select clusters of consumers based on where they purchased food

Cluster	Food store	Restaurant	Carryout	Vending	Cafeteria ¹
Home Cookers	93.1	2.5	1.1	0.3	0.1
Working Family	69.6	3.3	22.0	0.3	0.2
Carryout	34.8	3.7	57.3	0.1	0.2
Young Professional	33.8	8.2	40.4	14.2	0.8
High Service	46.8	42.8	5.0	0.4	0.4
Office	72.6	4.2	3.8	0.7	14.7
Manager	52.7	7.3	4.3	1.0	28.1
City Office	27.9	7.0	7.3	2.6	52.8
Students and Faculty	54.2	8.3	6.8	1.1	25.0

¹Both school and non-school cafeterias are combined.

Notes: Bold numbers identify the behavior around which a cluster was formed.

Totals do not add to 100, because not all sources of food are shown.

appear to play a much more important role for most consumers. A second observation is that both carryout restaurants and cafeterias have more than one cluster purchasing foods (g) from them, indicating major differences between the customers using these point-of-purchase sources. Three clusters formed around carryout food: Working Family, Carryout, and Young Professional. There are also differences in the shopping patterns, especially in the amount of food obtained from carryout restaurants, 22 to 57 percent. In addition, the Young Professional cluster is the only cluster discussed in this paper with a relatively high use of vending machines (14 percent). Similarly, four clusters formed around cafeterias as a source of food. The Office, Manager, and City Office clusters formed around non-school cafeterias, while the Students and Faculty cluster formed around school cafeterias (breakdown not shown). Except for City Office, these clusters all get at least half of the remaining food from stores, and make use of restaurants and carryout restaurants, though in different proportions.

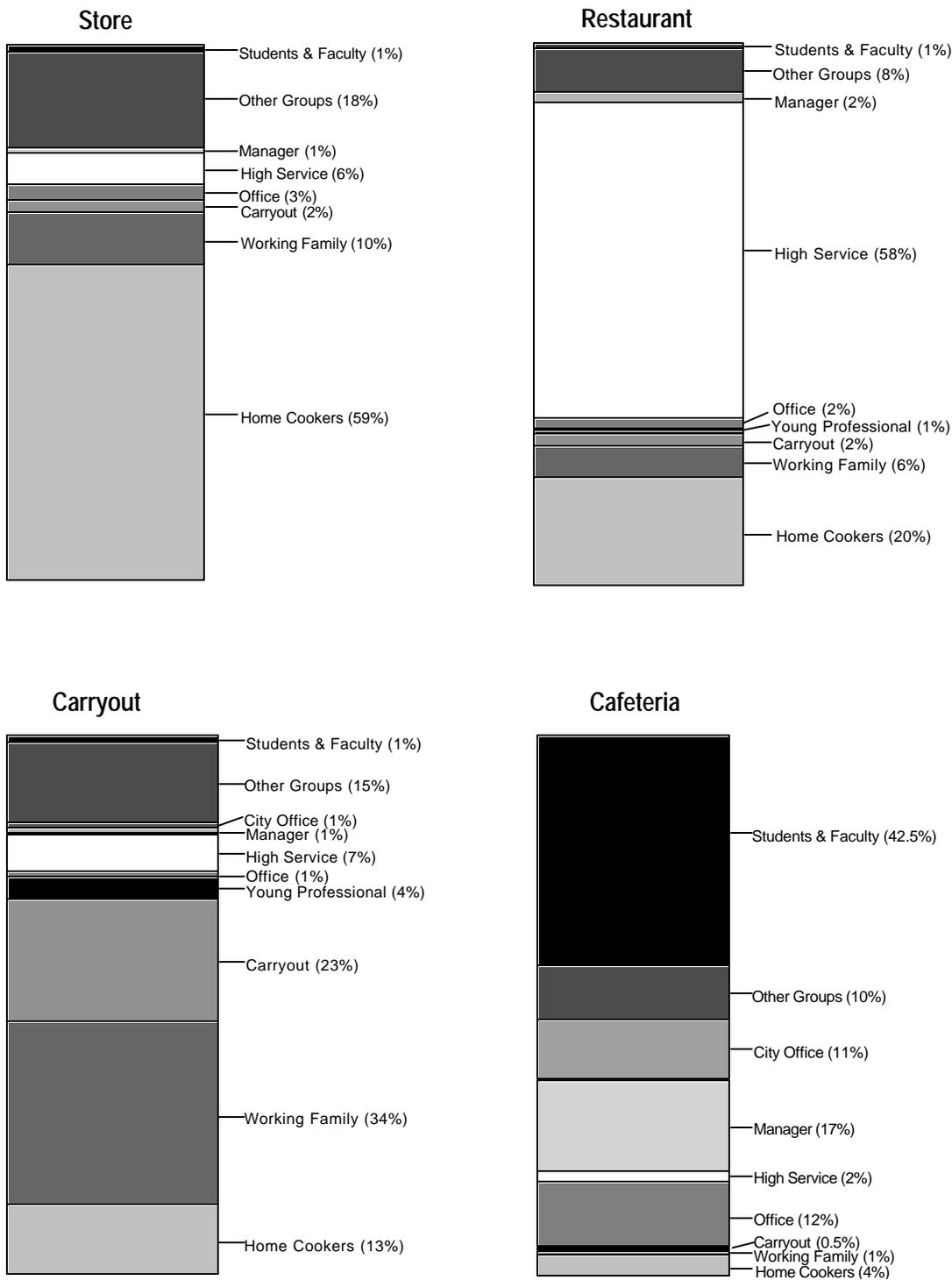
Market Profiles

When we examined consumption within markets (e.g., stores), we found that Home Cookers, the largest cluster, consumed 59 percent of all food (g) obtained from stores (fig. 1). The next two biggest clusters, Working Family and High Service, consumed 10 and 6 percent, respectively, of all food obtained from this source. This pattern of larger clusters representing larger portions of this market continued. “Other Groups” are clusters that formed but are not discussed in this paper. Each of these clusters in “Other Groups” had fewer than 100 observations; thus, statistical analysis may be misleading.

For restaurants, carryout restaurants, and cafeterias, the largest market share belonged to the cluster or clusters which formed around that source. For example, the High Service cluster, which formed around restaurants, represented 58 percent of the restaurant’s market share. For carryout restaurants, the Working Family, Carryout, and Young Professional clusters consumed over three-fifths (61 percent) of all food obtained from that market. Whereas

... Home Cookers, the largest cluster, consumed 59 percent of all food (g) obtained from stores.

Figure 1. Percentage of food consumed from selected sources, by cluster



Note: "Other groups" are clusters with fewer than 100 observations; these clusters are not discussed in this paper but are needed to complete the market profile.

the High Service cluster consumed 7 percent of the food in this market, the Young Professional cluster consumed less, 4 percent. However, the High Service cluster is a much larger cluster.

For the Carryout market, 70 percent of all food obtained here was consumed by three clusters: Working Family (34 percent of the grams of food consumed), Carryout (23 percent), and Home Cookers (13 percent). As expected, the Students and Faculty, Managers, Office, and City Office clusters consumed 83 percent of the food in the school and non-school cafeteria market. No other cluster consumes a large part of their food from this source, indicating the cafeteria market is fairly focused on these four clusters.

Conclusion

Americans who report in detail what food they eat, where they eat it, and where they buy it provide us with an alternative picture of food consumption based on the quantity of food (g) consumed. This varies from the more common picture based on food expenditures and sales. While it is true that Americans obtain food from many retail and home-grown sources, 75 percent of the adult population purchased over half of their food measured in grams from retail food stores. Thus we have a very different picture from the one presented by the use of food expenditure data. This alternative picture allowed us to ask two questions, what are the unique characteristics of people who shop for food in different establishments, and how can this information be used by these establishments and by public policymakers?

An examination of the data to determine the importance of each cluster to each type of retail vendor shows that,

among the people in our sample, Home Cookers purchase 59 percent of all the grams of food that were sold in retail stores, 20 percent of restaurant food, and 13 percent of the food from carryout establishments. The clusters most likely to be consumers of carryout food were the Young Professional, Working Family, and Carryout. People in these groups tend to be younger, employed, and have some college education.

Policymakers can use this information to determine how policies will affect different market segments: stores, restaurants, cafeterias, or carryout establishments. Owners and marketers of these establishments can determine where else their customers are obtaining food and design an appropriate marketing strategy.

Future research needs to address the effect that the choice of where to obtain food has on the quality and healthfulness of the diet. Identifying the consumers who are the first to make changes to their shopping habits, as well as identifying their preferences, will help retailers and those who design public food policy to serve consumers better.

References

- Carlson, A., Kinsey, J., & Nadav, C. (1998). *Who Eats What, When and From Where?* Minneapolis, MN: The Retail Food Industry Center, University of Minnesota. Working Paper Series.
- Food Institute. (1997). U.S. Food Service Industry Segments. *Food Institute Review*, 44, 3.
- Hartigan, J.A. (1985). Statistical Theory in Clustering. *Journal of Classification*, 2, 63-76.
- Kinsey, J. (1983). Working wives and the marginal propensity to consume food away from home. *American Journal of Agricultural Economics*, 65, 10-19.
- MacQueen, J.R. (1967). *Some Methods for Classification and Analysis of Multivariate Observations*. Paper presented at the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA.
- Prochaska, F., & Shrimper, R. (1973). Opportunity cost of time and other socioeconomic effects on away-from-home-food consumption. *American Journal of Agricultural Economics*, 66, 595-603.
- Putnam, J.J., & Allshouse, J.E. (1996). *Food Consumption, Prices, and Expenditures, 1970-94*. U.S. Department of Agriculture, Economic Research Service. Statistical Bulletin No. 928.
- SAS Institute, Inc. (1989). *SAS/STAT User's Guide, Version 6* (4th ed.). Cary, NC: SAS Institute, Inc.
- Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.
- U.S. Department of Agriculture, Agricultural Research Service. (1994). *Continuing Survey of Food Intakes by Individuals (CSFII)*.