REPLICATION OF EATING PATTERNS

John L. Stanton

The use of eating patterns as a way of describing the dietary habits of the American population is becoming more ubiquitous. However, to best utilize the concept of eating patterns in the context of large scale dietary surveys like NFCS, a number of critical issues must be addressed by the researchers. The resolution of these issues is significant since different decisions can lead to very different eating patterns being identified.

The three issues, addressed in this paper are the unit of measurement of dietary intake, level of aggregation of food groups and the statistical methods used to cluster individual diets into discrete patterns. Before one tries to resolve the above issues it should be pointed out that the primary source for the answer lies within a well defined statement as to what the researcher means by eating patterns. Many of the decisions one must make with respect to analytic alternatives are influenced by the end objective of the researcher. Therefore although the discussion from this point addresses analytic consideration one should not demean the significance of a well defined initial concept of eating patterns.

With respect to the unit of analysis three variables are most often considered. Food frequency, gram intake and caloric intake. By food frequency one means the binomial response of "ate" or "didn't eat" a food item. If a person had soup once in a day then a 1 would be his/her response for the soup item, or if a person had milk three times during the day the binomial response can be aggregated to a 3. The primary instance for using food frequency is when the researcher is interested in whether foods appear together (covary) in a day (or whatever unit of time) rather than the amount of intake. A difficulty with using frequency is that rather different dietary habits can be recorded as similar. For example if one takes milk and sugar in coffee each day and another puts milk and sugar on cereal each day, the frequency record for those two individuals would be identical even though the cereal consumer would have ingested much more milk and sugar than the other. Using grams or calories overcomes the above problem by indicating how much of each item is ingested as well as whether an item is ingested.

Also using grams or calories other than frequency permits the use of more powerful statistical methods like Pearson product moment correlations which should not be applied to ordinal or nominal data.

The second issue to be addressed is what level of food aggregation to use. It is almost impossible to treat each food item as a separate food for determining eating patterns. Not only is it impractical but it also may give very misleading results through loss of information due to over-disaggregation. By aggregation one means creating a larger defined food group, and aggregating all grams, calories or frequency together for any items consumed with that food group. Therefore if a person ate 100 grams of milk and 50 grams of cheese, a dairy group would have 150 grams, or if frequency was used the value for dairy would be 2. A number of different aggregations have been used and presented in the literature. Most commonly used are the 15 groups in Ten State Nutrition Survey, 18 groups used in the

Health and Nutritional Examination Survey, and the 48 groups for the Nationwide Food Consumption Survey using the first two digits of the food item codes. The actual choice of aggregates does have an impact on final patterns exacted. Smaller food aggregates often yield patterns which look more like meal components, like bread and butter/margarine, or soup and crackers. These are specific but may not give much of a picture of a daily pattern. Larger aggregates give a more descriptive picture of a daily pattern; e.g., meat, vegetables/fruits and desserts but little insight into what was specifically eaten. The decision again rests on the researchers' requirements.

The next issue to address is how will the foods be grouped in patterns, e.g., what statistical models will be used? There are at least two common methods applied to data of this sort. One method would be hierarchial clustering analysis. There are a number of these clustering programs available at most well equipped computer centers. Examples would be H-group in Veldman and PIM through P3m in BMDP. These models in general presume each case or variable is a single group and combine the groups into larger groups based on some amalgamation rule. At each step the two most similar clusters are joined to form a new cluster until a single cluster is obtained that includes all cases. However a shortcoming is the large time and space requirements. Therefore the size of the problem is often limited when using large government studies. For example, an estimate would be if you used 20 food groups in the analysis, space limitations in most computer systems prevent the handling over 370 cases. This is an obvious shortcoming when most studies have thousands of cases. An example where the analysis has been used in food and nutrition is "Determinants of Food Usage Behavior: A Market Segmentation Approach," published by the Marketing Science Institute, where 700 respondents and 10 variables were clustered using this algorithm.

The second method, which is more frequently used, is factor analysis. The factor analytic technique enables one to see whether some underlying pattern of relationships exists such that the data may be "rearranged" or reduced to a smaller set of factors or components. These may be taken as the source variables accounting for the observed interrelations in the data. Specifically related to food and nutrition, the source variables are the patterns of consumption which drive or motivate the choice process in the selection of foods. That is, even through a person may choose a wide variety of items on any day, the commonality of the items can be extracted and identified from the data. Therefore a person who eats milk with cereals and fruit juice each morning may be part of a "breakfast eating pattern." The juices and cereals may vary but if the person eats disproportionately more of these common items than the population, one might extract this breakfast group.

Factor analysis is based on the hypothesis that the observed relationships between variables (food and/or nutrients) are the results of some underlying regularity in the data, (eating pattern) rather than a stochastic choice model.

The advantage of factor analysis is that the basic data unit is the correlation matrix and therefore is not restricted by the number of cases. It does, however, in general, have a restriction on the number of variables.

For reasonably large computer systems you may have a maximum of about one hundred variables. Therefore if one used the 2 digit NFCS food codes (48) it would be no problem on most systems.

A disadvantage with factor analysis is that it is not a unitary concept and it subsumes a fairly large variety of procedures. One can not in this paper discuss all the issues in factor analysis but clearly a working knowledge of the key issues is necessary to choose among the variety of factor procedures.

The above discussion points out three critical issues in the analysis of eating patterns. It is by no means an exhaustive list of issues. The researcher must rely on the basic conceptual framework which drives the analysis to answer these three as well as the other critical issues in eating pattern analysis.