Nutrient Variability and Reliability: What to put in a food table?

William M. Rand+
Jean A.T. Pennington*

This presentation addresses the question of nutrient variability from the point of view of what data should be in a food composition table. The starting point is that the user needs information on both placement and spread of each nutrient in each food, and thus the fundamental question is: Given a set of data consisting of replicate measurements of a single nutrient measured in different samples of a single type of food, how should these data be summarized in a few useful statistics?

It needs to be recognized that there are many different reasons why such a set of replicate measurements will not all be identical. Differences will arise because the samples differ genetically and in the environmental conditions under which they developed; after harvest or slaughter, variability is introduced by differences in processing, preparation, and preservation; the samples themselves will differ in how they are prepared for analysis and how they are actually analyzed; and finally, the "food" entry itself may include different items. However, from the perspective of the user of a given food composition table looking at the entry for a particular food, the source of variability is irrelevant. Of interest is only the fact that these "replicates" represent a distribution of what the consumer may consume, and the food table users needs guidance on that point.

Currently each major food composition database has its own conventions for the data it includes, and each presents its data slightly differently. In general, data are included which indicate the nutrient levels most likely to be encountered (the mean, median or mode -- the middle of the distribution), how variable the nutrient is (the standard deviation, the standard error, the range), and how many separate or independent observations these summary statistics are based on. From these statistics, the user estimates typical consumption (from measures of the middle of the distribution) and risks of toxicity and undernutrition (from measures of variability).

Given a set of data, there are a number of well-defined statistics that permit us to estimate where the middle and extremes of a variable are likely to be; however, to interpret anything but percentiles interpolated directly from the data (e.g., median and 5th and 95th percentiles), it is necessary to assume a specific probability distribution (such as normal or lognormal) for each nutrient in each food. This is a strong assumption which allows us to work more confidently with small data sets. The aim of this preliminary study is to explore the basis of making distributional assumptions, and the implications of these assumptions in terms of data presentation in food composition databases. We examined data from nutrient analyses of replicate measurements of more than 200 different foods collected in the Food and Drug Administration's (FDA's) Total Diet Study from 1982 to 1989. We present here our results for two nutrients, sodium and calcium.

The usual distributions that are assumed by workers using food composition data are the normal and the lognormal, and, at one level, skewness distinguishes between these. If a distribution of nutrient content is not skewed, it can often be assumed to be normally distributed, and if the distribution of the log of the nutrient is not skewed, it is assumed to be lognormally distributed. (Throughout, we work with skewness that is normalized, so that a value smaller than -1.96 or larger than +1.96 suggests 0.05 significance.)

Our first observation, illustrated in Figure 1, is that a full range of distributional shapes exists. Note that those that are not skewed could be normal, those skewed positively could be lognormally distributed and those skewed negatively could be neither normal nor lognormal. These are not definitive results since they were selected to illustrate this point from a large number of different foods. These particular histograms may well represent the extremes of inherent sampling variability; however, they do show that categorical assumption of normality or lognormality, for at least sodium and calcium, does not rest on a firm basis.

We next hypothesize that shape of nutrient distribution is related to type of food, and that similar foods will have similar distributions. Figure 2 shows skewness plotted against the food groupings that are used in FDA's Total Diet Study (arbitrarily assigned the first 31 integers). This shows that, contrary to our expectations, food grouping is not closely related to the skewness of the distribution.

Finally, we compared, for each food, the skewness of the nutrient (to see if it might be normally distributed) with the skewness of the log transform of that nutrient (to see if it might be lognormally distributed). Figure 3 shows the results for the two nutrients, sodium and calcium. Based on the skewness, in almost half of the foods looked at, the nutrient could be **either** normal **or** lognormal while in about a sixth of the foods these nutrients could be **neither** normal **nor** lognormal.

This preliminary data exploration leads us to the following conclusions and recommendations:

1. We cannot specify with any certainty the distribution of any nutrient in any food; therefore, food composition databases should, as a default, list interpolation estimates of the 5th, 50th (median) and 95th percentiles.

2. The distribution of nutrient in a food appears uncorrelated with the type of food; therefore, it is likely that distribution shape may result from sources of variability other than genetics. Data currently exist which contain information on this and should be analyzed.

+ William Rand
Tufts University
Boston, Massachusetts

* Jean Pennington
Food and Drug Administration
Washington, District of Columbia

FIGURE 1: SELECTED DATA DISTRIBUTIONS

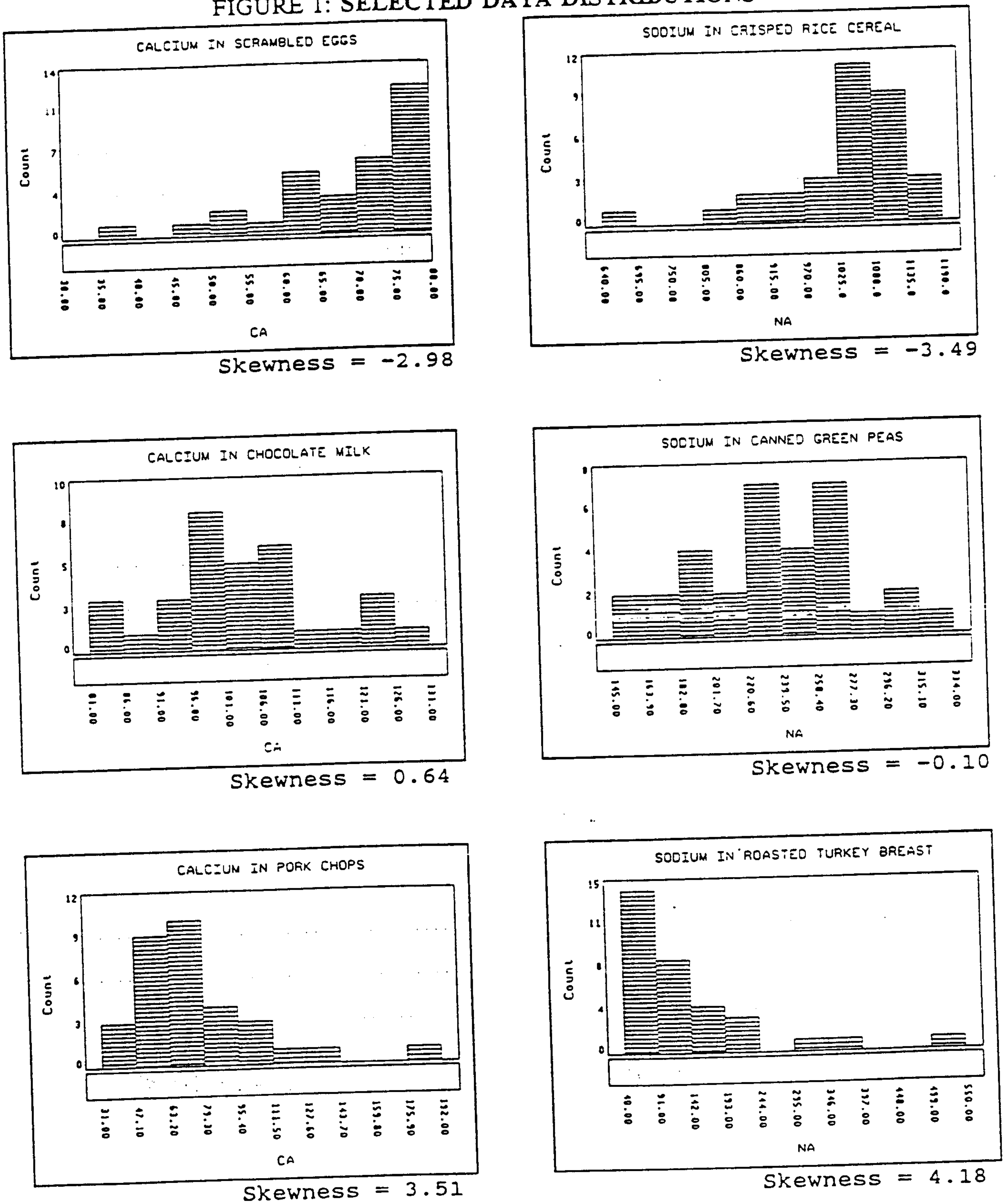


FIGURE 2: SKEWNESS VERSUS FOOD GROUP

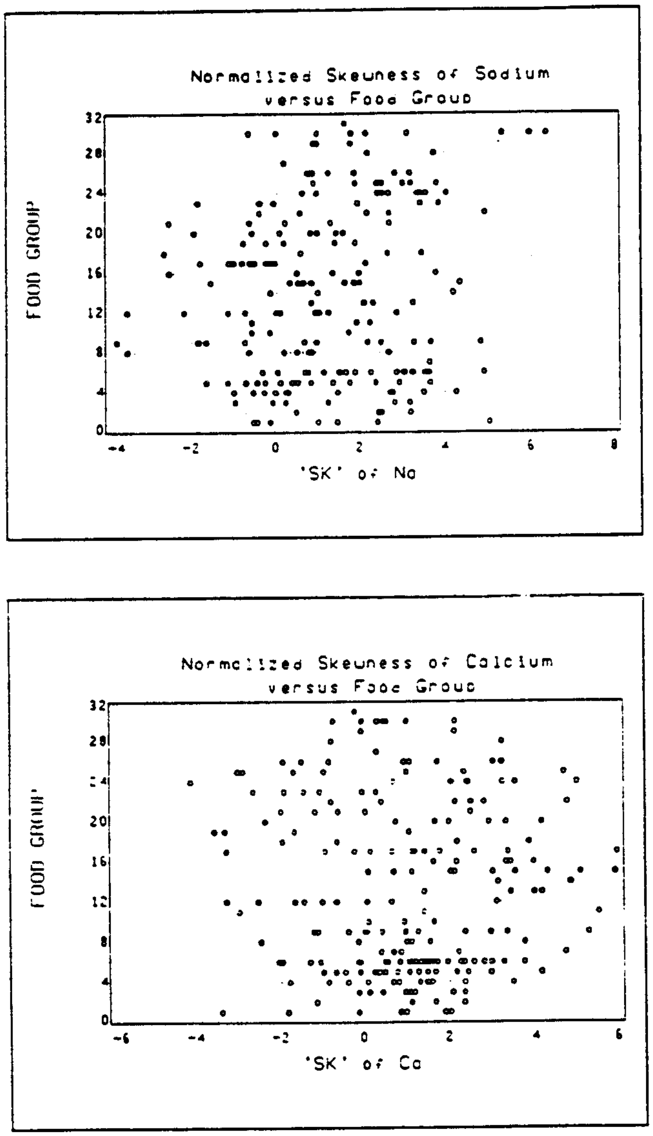
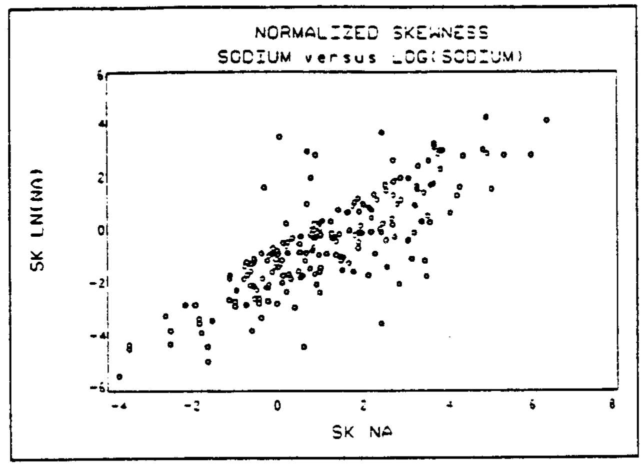
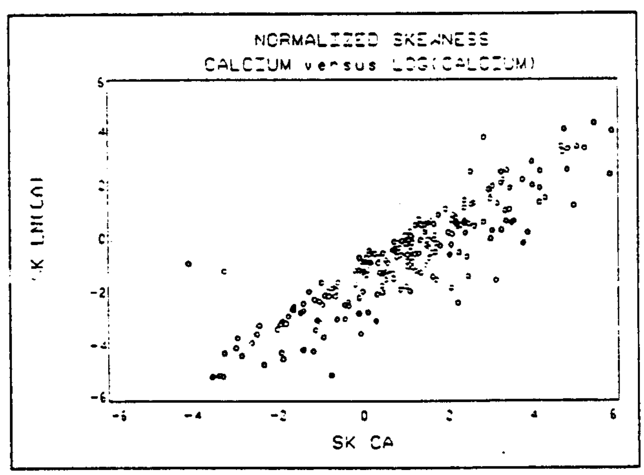


FIGURE 3: SKEWNESS VERSUS LOG SKEWNESS



Sodium Distribution	No. of Foods	Percent
Either Normal or Log-Normal	95	47.0%
Normal but not Log-Normal	34	16.8
Log-Normal but not Normal	45	22.3
Neither Normal nor Log-Normal	28	13.9
Total Number of Foods Examined	202	



Calcium Distribution	No. of Foods	Percent
Either Normal or Log-Normal	102	45.7%
Normal but not Log-Normal	36	16.1
Log-Normal but not Normal	50	22.4
Neither Normal nor Log-Normal	35	15.7
Total Number of Foods Examined	223	