Nutrient Variability

Jean AT Pennington, PhD, RD[1], Richard H Albert, PhD[1], and William M Rand, PhD[2]

Measures of nutrient variability are particularly useful to database compilers, food analysts, and food manufacturers. Database compilers use them to make decisions regarding the aggregation and compilation of data from various sources, food analysts use them to gauge the potential results of prospective analyses, and food manufacturers use them in product development and nutrition labeling. Measures of nutrient variability are also of importance to dietitians, nutritionists, and researchers who use food composition databases to plan and evaluate the diets of patients, clients, and study participants.

Causes of nutrient variability include inherent, environmental, and processing factors such as variety, agricultural conditions, storage conditions, and cooking methods. The causes of nutrient variability are food specific and are, for the most part, difficult to quantitate or separate. Some nutrient variability is the result of artifacts such as sampling scheme, analytical methods, quality control, laboratory bias, and statistical treatment of data.

Information about nutrient variability is usually expressed as a standard deviation (SD), standard error (SE), or coefficient of variation (CV). A comparison of mean and median values and an examination of the range of values also provide some indication of how a nutrient varies in a food. However, nutrient variation is most visually apparent with the use of frequency distributions. The examples of nutrient variation presented here are from the Food and Drug Administration's (FDA) Total Diet Studies, 1982-91.

Figure 1.
CALCIUM IN WHOLE FLUID MILK

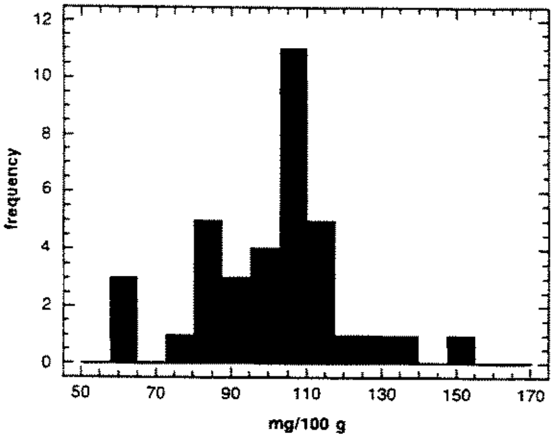
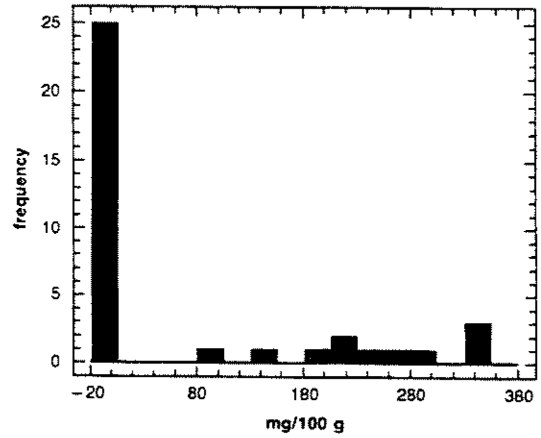


Figure 2.
SODIUM IN COOKED WHITE RICE

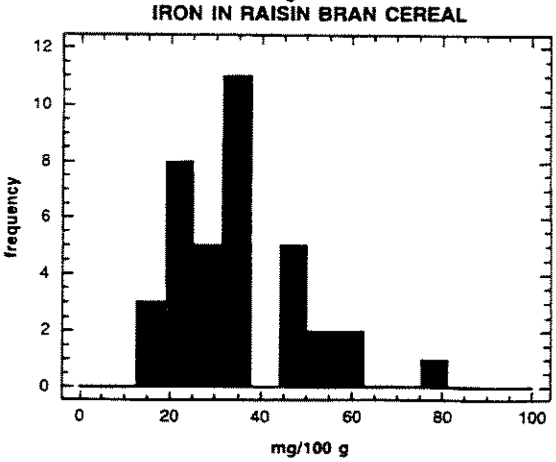


The level of calcium in whole milk (Figure 1) might vary because of environmental factors such as season and the type of diet fed to cows and because of factors inherent in the cows such as stage of lactation, age, and species. Processing variables include the pooling of milk from many cows within a dairy and the pooling of milk from various dairies before it is processed, packaged, and sold to consumers. Milk, which is commonly recommended as a source of calcium, provides an average level of 247 ± 50 mg of calcium per eight fluid ounces, however, the amount of calcium varies from 142 to 378 mg per eight fluid ounces.

The variability of sodium in processed foods is primarily dependent upon the quantity of salt and other sodium-containing compounds added by the manufacturer. The sodium content of processed foods is often brand specific. The sodium content of various brands of condensed, diluted vegetable beef soup from the FDA Total Diet Study is 669 ± 226 mg/cup with a range of 284 to 1,624 mg/cup. This variation indicates the importance of brand- specific information for the nutrient content of some foods, especially if the nutrient content of individual diets is to be assessed.

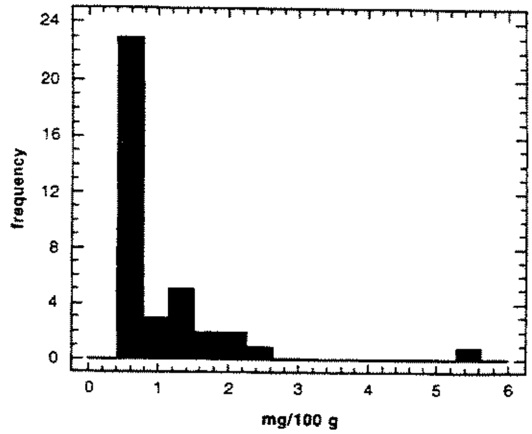
The mean and median values for the sodium content of cheddar cheese, 170 and 168 mg/ounce, respectively, are similar; however, the range (113 to 249 mg/ounce) indicates the different levels of sodium added to cheese during processing.

Figure 3.
IRON IN RAISIN BRAN CEREAL



Frequency distributions for sodium in cooked white rice, oatmeal, farina, and corn grits are similar, but not normal (See Figure 2 as an example). These distributions show a high frequency of low values and a scattering of values at the higher end. The "regular" form of these four grain products was supposed to have been purchased, but sometimes instant products were mistakenly collected by the FDA inspectors. The instant products usually contain added sodium compounds and are higher in sodium than the regular products. Another source of sodium variation has been salt added to these products at the contract kitchen where the foods are prepared and cooked. Some of the sodium values from the earlier years of the studies are higher because

Figure 4.
IRON IN CANNED TOMATO SAUCE



the food preparers added salt according to package directions. The food preparers were subsequently instructed not to add salt to these foods during cooking. Since then the sodium values for these foods have become consistently low. The mode more accurately reflects the sodium content of these foods than the mean; the mean value reflects neither the salted nor the unsalted product.

Iron in raisin bran cereal (Figure 3) illustrates variability caused by different fortification levels. This product is commonly sold unfortified, fortified at a 25% U.S. RDA for iron, or fortified at a 100% U.S. RDA for iron. The mean value of iron in this product (0.96 ± 0.14 mg/ounce; range 0.73-1.46 mg/ounce) does not accurately reflect any of the subtypes of which it is composed. If a brand-specific product had been collected and analyzed, the range and standard deviation would have been smaller.

Frequency distributions are useful for identifying outliers. The outlying high iron level in canned tomato sauce shown in Figure 4 may have resulted from iron contact or contamination at the manufacturing plant, but it does not reflect the usual iron content of this product. Outliers may result from errors in the purchase, compositing, or analysis of a food. If the nutrient values are off by a factor of 10, 100, or more, there may have been errors in sample dilution or in data entry. To the extent possible, the identification of outliers should be confirmed by analysts in the laboratory. If there are no reserve samples to reanalyze and no other laboratory information to rely upon, the data evaluator must make the final decision about which values to include in a database. If there are sufficient numbers of samples, outliers may have no effect on mean values. However, when they do, it is best to exclude them or to use median values. Decisions regarding the treatment of outliers should be documented for later reference.

Most users of databases want values that reflect the value most likely to occur or the value with the highest probability of occurring. In most cases, this would be the mean or median. If there are outliers, the median might be a better choice. In some cases (as with the sodium in white rice), the mode (most frequently occurring value) might be the best solution. The data compiler must be knowledgeable about each food and the variables that affect nutrient values.

When food composition data are published or otherwise made available, it is desirable that the data be accompanied by complete food descriptions, sampling design, number of samples, analytical method, quality control information, and median and mean values with an estimate of variance (or individual data points). Additional information that might account for unusual levels of nutrients should be provided. This might include information about food additives (e.g., magnesium additives in canned green beans), fortification (e.g., iron in ready-to-eat cereals), or processing (e.g., mechanical deboning of meat which increases calcium content). Information to explain large variances should also be provided. For example, a large standard deviation for vitamin A in sweet potatoes or vitamin C in grapefruit may be the result of several cultivars in a sample.

When evaluating frequency distributions, one might consider the following questions:

1. Was the sampling design appropriate? If not, how might it have altered the frequency distribution?
2. If the frequency distribution is not normal (i.e., non-Gaussian), is it skewed or bi-modal? Are there apparent explanations for the non-normal distribution (e.g., several different populations)?
3. Are outliers apparent and, if so, should they be omitted?
4. Does the distribution suggest that the mean, median, or mode might be the best choice for the "typical" level of the nutrient in the food?
5. Is the concentration of this nutrient in this food of any practical importance? If a food is not a good source of a nutrient (e.g., copper in milk), the distribution will probably not be normal because there may be many zero, trace, or very low values. If a food is not a potential source of a nutrient (e.g., less than 2% of the Daily Value), the shape of the frequency distribution is of little practical consequence.

Dietary advice given by dietitians and nutritionists should be based on realistic information about food composition. Specific foods that are promoted as being good sources of a nutrient should also be reliable sources of that nutrient. How much can a nutrient vary before the "typical" value is considered unreliable? Might health professionals provide better dietary advice if the databases they use have information about nutrient variability?

Questions of interest with regard to nutrient variation are still rather basic:

1. How variable are nutrients in foods?
2. Are some nutrients more variable than others?
3. Are there similarities in variance among food groups?
4. Are there similarities in variance among nutrients?
5. What are the criteria for using means, medians, and modes?
6. Does the use of medians vs. means vs. modes affect dietary assessments?
7. What happens to variability as more data are collected?

Preliminary results from the Total Diet Study indicate that for minerals, there are no apparent similarities in variation among food groups and that the use of medians (instead of means) has little effect on assessment of daily nutrient intakes.

The more samples that are analyzed, the more valid the database becomes and the more clearly the outliers are identified. However, more data do not decrease variability. More data allow variability to be more clearly defined. With more data, one can put more confidence in median or mean values, but this does not allow for clearer predictions of the nutrient levels of any specific sample.

[1] Center for Food Safety and Applied Nutrition, Food and Drug Administration, Washington, DC.
[2] Department of Community Health, Tufts University School of Medicine, Boston, MD.