

# StatLib---1993 Graphics Exposition

## "Serial Correlation or Cereal Correlation ??"

Call for Poster Presentations for the 1993 Statistical Graphics Exposition

### REVISED README FILE

(new breakfast cereal data and new information about the data)

Every two years the Section on Statistical Graphics sponsors a special exposition where one or more data sets are made available, analyzed by anyone interested and presented in a special poster session at the Annual Meeting.

For the 1993 Statistical Graphics Exposition, there are two datasets to analyze, one synthesized, one real:

**OSCILLATOR TIME SERIES** - a synthesized univariate time series with 1024 observations. These data are similar to those which might be found in a university or industrial laboratory setting, or possibly from a process monitor on a plant floor. They show obvious structure, but there is more than one feature present, and good graphics are key to uncovering the features. The objective is to find ALL the features. At the Exposition next year, the algorithm and coefficients by which the dataset was constructed will be presented, along with the stages of analysis which would uncover the features. Some questions to consider:

- What graphics are helpful in selecting the right analytical tools?
- What combinations of graphics are essential to finding all the features?
- For what features are the traditional graphics and analytical tools weak?
- Are there graphics that you can retrospectively develop which more clearly reveal the features which were hard to uncover?

The [oscillator](#) data are available in an ASCII file, one observation per record. To obtain the data, send an email message to [statlib@lib.stat.cmu.edu](mailto:statlib@lib.stat.cmu.edu) containing the single line:

```
send oscillator from 1993.expo
```

**BREAKFAST CEREAL DATA (REVISED)**- a multivariate dataset describing seventy-seven commonly available breakfast cereals, based on the information now available on the newly-mandated F&DA food label. What are you getting when you eat a bowl of cereal? Can you get a lot of fiber without a lot of calories? Can you describe what cereals are displayed on high, low, and middle shelves? The good news is that none of the cereals for which we collected data had any cholesterol, and manufacturers rarely use artificial sweeteners and colors, nowadays. However, there is still a lot of data for the consumer to understand while choosing a good breakfast cereal.

Two new variables have been added to the data (end of each record):

```
weight (in ounces) of one serving (serving size) [weight]  
cups per serving [cups]
```

Otherwise, the data are the same, except for minor typo corrections. The addition of these variables (suggested by Abbe Herzig of Consumers Union. Cereals vary considerably in their densities and listed serving sizes. Thus, the serving sizes listed on cereal labels (in weight units) translate into different amounts of nutrients in your bowl. Most people simply fill a cereal bowl (resulting in constant volume, but not weight). The new variables help standardize other ways, which provides other ways to differentiate and group cereals.

Here are some facts about nutrition that might help you in your analysis. Nutritional recommendations are drawn from the references at the end of this document:

- Adults should consume between 20 and 35 grams of dietary fiber per day.
- The recommended daily intake (RDI) for calories is 2200 for women and 2900 for men.
- Calories come in three food components. There are 9 calories per gram of fat, and 4 calories per gram of carbohydrate and protein.
- Overall, in your diet, no more than 10% of your calories should be consumed from simple carbohydrates (sugars), and no more than 30% should come from fat. The RDI of protein is 50 grams for women and 63 grams for men. The balance of calories should be consumed in the form of complex carbohydrates (starches).
- The average adult with no defined risk factors or other dietary restrictions should consume between 1800 and 2400 mg of sodium per day.
- The type and amount of milk added to cereal can make a significant difference in the fat and protein content of your breakfast.

One possible task is to develop a graphic that would allow the consumer to quickly compare a particular cereal to other possible choices. Some additional questions to consider, and try to answer with effective graphics:

- Can you find the correlations you might expect? Are there any surprising correlations?
- What is the true "dimensionality" of the data?
- Are there any cereals which are virtually identical?
- Is there any way to discriminate among the major manufacturers by cereal characteristics, or do they each have a "balanced portfolio" of cereals?
- Do the nutritional claims made in cereal advertisements stand the scrutiny of data analysis?
- Are there cereals which are clearly nutritionally superior, or inferior? Are there clusters of cereals?
- Is a ranking or scoring scheme possible or reasonable, and if so, are there cereals which are nutritionally superior or inferior under all reasonable weighting schemes?

The variables of the dataset are listed below, in order. For convenience, we suggest that you use the variable name supplied in square brackets.

Breakfast cereal variables:  
 cereal name [name]  
 manufacturer (e.g., Kellogg's) [mfr]  
 type (cold/hot) [type]  
 calories (number) [calories]  
 protein(g) [protein]  
 fat(g) [fat]  
 sodium(mg) [sodium]  
 dietary fiber(g) [fiber]

```
complex carbohydrates(g) [carbo]
sugars(g) [sugars]
display shelf (1, 2, or 3, counting from the floor) [shelf]
potassium(mg) [potass]
vitamins & minerals (0, 25, or 100, respectively indicating
'none added'; 'enriched, often to 25% FDA recommended'; '100% of
FDA recommended') [vitamins]
weight (in ounces) of one serving (serving size) [weight]
cups per serving [cups]
```

Manufacturers are represented by their first initial: A=American Home Food Products, G=General Mills, K=Kellogg, N=Nabisco, P=Post, Q=Quaker Oats, R=Ralston Purina)

The breakfast [cereal](#) data are available in an ASCII file, one cereal per record, with underscores in place of the spaces in the cereal name, and spaces separating the different variables. The value -1 indicates missing data. To obtain the data, send an email message to: [statlib@lib.stat.cmu.edu](mailto:statlib@lib.stat.cmu.edu) containing the single line:

```
send cereal from 1993.expo
```

Work alone or put together a team of data analysts to look at one or both of these two data sets! Try to answer the questions posed here or conduct an exploratory analysis to find and answer your own questions.

To participate in the Exposition, you must submit a contributed paper abstract for inclusion in the formal ASA Contributed Paper Program. This reserves a poster session slot for you. Your abstract, on the official ASA abstract form, is due by the contributed paper deadline, February 1, 1993.

If you do not have electronic mail access, try to get the data files from someone who already has them. If you cannot obtain the data via electronic mail, contact David Coleman, AMCT-D, Alcoa Technology Center, Alcoa Center, PA 15069, or e-mail [COLEMAN1@ncf.al.alcoa.com](mailto:COLEMAN1@ncf.al.alcoa.com)

## References:

National Research Council, 1989a. "Diet and Health: Implications for Reducing Chronic Disease Risk". National Academy Press, Washington, D.C.

National Research Council, 1989b. "Recommended Dietary Allowances, 10th Ed." National Academy Press, Washington, D.C.

National Cancer Institute, 1987. "Diet, Nutrition, and Cancer Prevention: A Guide to Food Choices," NIH Publ. No. 87-2878. National Institutes of Health, Public Health Service, U.S. Department of Health and Human Service, U.S. Government Printing Office, Washington, D.C.