

Stat 445: Midterm Project #1

“Part of This Nutritious Breakfast!”

Due: Wednesday, February 23 *in class*

Dataset Description: The dataset for this project is available at:

<http://unixlab.stat.ubc.ca/~stat445/projects/>

It contains per-serving nutritional information and grocery shelf location for 77 breakfast cereals from 7 manufacturers. The variables have the following meanings:

name	name of cereal
mfr	manufacturer (A = <i>American Home Food Products</i> ; G = <i>General Mills</i> ; K = <i>Kelloggs</i> ; N = <i>Nabisco</i> ; P = <i>Post</i> ; Q = <i>Quaker Oats</i> ; R = <i>Ralston Purina</i>)
type	C = <i>cold</i> or H = <i>hot</i>
calories	calories per serving
protein	protein (grams)
fat	fat (grams)
sodium	sodium (milligrams)
fiber	dietary fiber (grams)
carbo	complex carbohydrates (grams)
sugars	sugars (grams)
potass	potassium (milligrams)
vitamins	typical percentage of Food and Drug Administration (FDA) recommended daily amount of vitamins and minerals
shelf	supermarket display shelf (counting from 1 = <i>floor</i>)
weight	recommended serving size (ounces)
cups	recommended serving size (cups)
rating	<i>Consumer Reports</i> rating ¹

Note: In the dataset, the special value -1 indicates a missing data item.

Project: Analyze this dataset and produce a report of your findings. Use any graphical and statistical techniques that seem appropriate to you.

Your report should be approximately 6 pages in length (maximum 8 pages) including any essential figures and tables, and these should be fully integrated into the text of the report. Your report should start with an “Executive Summary” of no more than one page (and included in the above page limit) providing a brief overview of your report including any key statistical findings and your interpretation of them.

¹The **rating** was calculated by *Consumer Reports* using an undisclosed formula presumably based, in part, on the nutritional content.

Additional Guidelines: The projects in this course are intended to serve at least three purposes:

- to provide experience in formulating problems statistically;
- to provide experience in statistical data analysis; and
- to provide experience in writing reports.

Therefore, your emphasis in these projects should be on statistical problem solving and on writing a clear, complete, and concise report.

You should write your report so that an individual without much statistical training, such as your future boss or a future client for whom you are acting as a statistical consultant, can understand your analysis and your conclusions. In particular, your report should make very clear how you have formulated the problem as a statistical problem and how the results of your analysis relate to making conclusions that would be relevant to someone interested in the original problem (and *not* someone interested in, say, linear regression theory or R programming techniques).

You will also want to discuss any reservations you have about the original data, the appropriateness of the analysis you carried out, and your conclusions.