R Markdown example with cereal data

Chris Bilder

# Cereal data example

The purpose of this document is to introduce some very basic aspects of using R Markdown.

## Basics

A few years ago, I collected information on the nutritional content of dry cereals at a grocery store. This was done by first noting that one side of one aisle in many grocery stores usually contains all the cereals within a store. For example, below shows what the cereal aisle used to look like in the HyVee at 5020 N 27th St. before its recent renovation (image file needs to be in same folder as .Rmd file):



My research hypothesis was that there were different mean nutritional contents by shelf. For example, lower shelves may have more sugar content cereals than higher shelves.

The data used for this example was collected from a store a few years ago (not the HyVee in the picture). Note that there were only four shelves at this store and my sample size was 10 from each shelf. Below is how I read the data into R.

> # Data file needs to be put in same folder as .Rmd file or
> # provide the full address to its location (either on your
> # computer or on the Internet)
> cereal <- read.csv(file = "http://www.chrisbilder.com/stat801/sections/2/cereal.csv")
> head(cereal)

 Cereal.ID Shelf Cereal size\_g sugar\_g fat\_g
1 1 1 Kellog's Razzle Dazzle Rice Crispies 28 10 0
2 2 1 Post Toasties Corn Flakes 28 2 0
3 3 1 Kellog's Corn Flakes 28 2 0
4 4 1 Food Club Toasted Oats 32 2 2
5 5 1 Frosted Cheerios 30 13 1
6 6 1 Food Club Frosted Flakes 31 11 0
 sodium\_mg
1 170
2 270
3 300
4 280
5 210
6 180

The shelves are numbered from lowest (1) to highest (4).

We need to adjust the nutritional content variables (sugar\_g, fat\_g, and sodium\_g) for the serving size because cereal boxes tend to have different serving sizes. Below is how I make the adjustment in R.

> cereal$sugar <- cereal$sugar\_g/cereal$size\_g
> cereal$fat <- cereal$fat\_g/cereal$size\_g
> cereal$sodium <- cereal$sodium\_mg/cereal$size\_g
> head(cereal)

 Cereal.ID Shelf Cereal size\_g sugar\_g fat\_g
1 1 1 Kellog's Razzle Dazzle Rice Crispies 28 10 0
2 2 1 Post Toasties Corn Flakes 28 2 0
3 3 1 Kellog's Corn Flakes 28 2 0
4 4 1 Food Club Toasted Oats 32 2 2
5 5 1 Frosted Cheerios 30 13 1
6 6 1 Food Club Frosted Flakes 31 11 0
 sodium\_mg sugar fat sodium
1 170 0.35714286 0.00000000 6.071429
2 270 0.07142857 0.00000000 9.642857
3 300 0.07142857 0.00000000 10.714286
4 280 0.06250000 0.06250000 8.750000
5 210 0.43333333 0.03333333 7.000000
6 180 0.35483871 0.00000000 5.806452

For example, the sugar value for Kellogg’s Razzle Dazzle Rice Crispies is $(sugar grams per serving)/(\# of grams per serving size)$ = 10/28 = 0.3571.

Alternatively, taking advantage of the knitr package, the calculation can be automatically done here using R:

$$(sugar grams per serving)/(\# of grams per serving size)=10/28=0.3571.$$

If these calculations were done outside of an equation, one could simply use something like: “The amount of sugar per gram is 0.3571”.