MATH 4044 – Statistics for Data Science

Practical Week 11

Question 1

The data file for this practical is called `cereals.sas7bdat` located in `mydata` library on the SAS OnDemand server. This data file contains nutritional information, rating (whether healthy or not) and supermarket shelf location for 77 breakfast cereals. Variables in that file are as follows:

Variable	Description
name	Name of cereal
mfr	Manufacturer of cereal where A = American Home Food Products; G = General Mills; K = Kelloggs; N = Nabisco; P = Post; Q = Quaker Oats; R = Ralston Purina
type	C = cold, H = hot
calories	Calories per serve
protein	Grams of protein
fat	Grams of fat
sodium	Milligrams of sodium
fiber	Grams of dietary fibre
carbo	Grams of complex carbohydrates
sugars	Grams of sugar
potass	Milligrams of potasium
vitamins	Vitamins and minerals, 0, 25, or 100, indicating the typical percentage of FDA recommended
shelf	Display shelf (1 = bottom, 2 = middle, or 3 = top, counting from the floor)
weight	Weight in ounces of one serving
cups	Number of cups in one serving
rating	Rating of the cereals calculated from Consumer Reports, out of 100. The higher the score, the healthier the cereal

(a) Is there a significant difference in ratings of cereals displayed on different shelves? Carry out a Kruskal-Wallis test and perform post-hoc tests if appropriate. Discuss the results.

(b) Convert the data to ranks and perform a one-way analysis of variance on ranks. Include post-hoc tests if appropriate. Discuss the results and compare to part (a).

Note: You can write your own code (see topic notes) or use Tasks. If you choose to use the Tasks menu, make sure that PROC GLM is used.