

W271 Assignment 2

Due Tuesday 23 June 2020 11:59pm

Instructions (Please Read Carefully):

- No page limit, but be reasonable
 - Do not modify fontsize, margin or line_spacing settings
 - This assignment needs to be completed individually; this is not a group project. Each student needs to submit their homework to the course github repo by the deadline; submission and revisions made after the deadline will not be graded
 - Answers should clearly explain your reasoning; do not simply ‘output dump’ the results of code without explanation
 - Submit two files:
 1. A pdf file that details your answers. Include all R code used to produce the answers. Do not suppress the codes in your pdf file
 2. The R markdown (Rmd) file used to produce the pdf file
- The assignment will not be graded unless **both** files are submitted
- Use the following file-naming convention:
 - StudentFirstNameLastName_HWNumber.fileExtension
 - For example, if the student’s name is Kyle Cartman for assignment 1, name your files follows:
 - * KyleCartman_assignment1.Rmd
 - * KyleCartman_assignment1.pdf
 - Although it sounds obvious, please write your name on page 1 of your pdf and Rmd files
 - For statistical methods that we cover in this course, use the R libraries and functions that are covered in this course. If you use libraries and functions for statistical modeling that we have not covered, you must provide an explanation of why such libraries and functions are used and reference the library documentation. For data wrangling and data visualization, you are free to use other libraries, such as dplyr, ggplot2, etc.
 - For mathematical formulae, type them in your R markdown file. Do not e.g. write them on a piece of paper, snap a photo, and use the image file.
 - Incorrectly following submission instructions results in deduction of grades
 - Students are expected to act with regard to UC Berkeley Academic Integrity

1. Strategic Placement of Products in Grocery Stores (5 points)

These questions are taken from Question 12 of chapter of the textbook.

In order to maximize sales, items within grocery stores are strategically placed to draw customer attention. This exercise examines one type of item—breakfast cereal. Typically, in large grocery stores, boxes of cereal are placed on sets of shelves located on one side of the aisle. By placing particular boxes of cereals on specific shelves, grocery stores may better attract customers to them. To investigate this further, a random sample of size 10 was taken from each of four shelves at a Dillons grocery store in Manhattan, KS. These data are given in the cereal_dillons.csv file. The response variable is the shelf number, which is numbered from bottom (1) to top (4), and the explanatory variables are the sugar, fat, and sodium content of the cereals.

1.1 (1 point): The explanatory variables need to be reformatted before proceeding further (sample code is provided in the textbook). First, divide each explanatory variable by its serving size to account for the different serving sizes among the cereals. Second, rescale each variable to be within 0 and 1. Construct side-by-side box plots with dot plots overlaid for each of the explanatory variables. Also, construct a parallel coordinates plot for the explanatory variables and the shelf number. Discuss whether possible content differences exist among the shelves.

1.2 (1 point): The response has values of 1, 2, 3, and 4. Explain under what setting would it be desirable to take into account ordinality, and whether you think that this setting occurs here. Then estimate a suitable multinomial regression model with linear forms of the sugar, fat, and sodium variables. Perform LRTs to examine the importance of each explanatory variable. Show that there are no significant interactions among the explanatory variables (including an interaction among all three variables).

1.3 (1 point): Kellogg's Apple Jacks (<http://www.applejacks.com>) is a cereal marketed toward children. For a serving size of 28 grams, its sugar content is 12 grams, fat content is 0.5 grams, and sodium content is 130 milligrams. Estimate the shelf probabilities for Apple Jacks.

1.4 (1 point): Construct a plot similar to Figure 3.3 where the estimated probability for a shelf is on the y -axis and the sugar content is on the x -axis. Use the mean overall fat and sodium content as the corresponding variable values in the model. Interpret the plot with respect to sugar content.

1.5 (1 point): Estimate odds ratios and calculate corresponding confidence intervals for each explanatory variable. Relate your interpretations back to the plots constructed for this exercise.

2. Alcohol, self-esteem and negative relationship interactions (5 points)

Read the example ‘**Alcohol Consumption**’ in chapter 4.2.2 of the textbook. This is based on a study in which moderate-to-heavy drinkers (defined as at least 12 alcoholic drinks/week for women, 15 for men) were recruited to keep a daily record of each drink that they consumed over a 30-day study period. Participants also completed a variety of rating scales covering daily events in their lives and items related to self-esteem. The data are given in the *DeHartSimplified.csv* data set. Questions 24-26 of chapter 3 of the textbook also relate to this data set and give definitions of its variables: the number of drinks consumed (**numall**), positive romantic-relationship events (**prel**), negative romantic-relationship events (**nrel**), age (**age**), trait (long-term) self-esteem (**rosn**), state (short-term) self-esteem (**state**).

The researchers stated the following hypothesis:

We hypothesized that negative interactions with romantic partners would be associated with alcohol consumption (and an increased desire to drink). We predicted that people with low trait self-esteem would drink more on days they experienced more negative relationship interactions compared with days during which they experienced fewer negative relationship interactions. The relation between drinking and negative relationship interactions should not be evident for individuals with high trait self-esteem.

2.1 (2 points): Conduct a thorough EDA of the data set, giving special attention to the relationships relevant to the researchers’ hypotheses. Address the reasons for limiting the study to observations from only one day.

2.2 (2 points): The researchers hypothesize that negative interactions with romantic partners would be associated with alcohol consumption and an increased desire to drink. Using appropriate models, evaluate the evidence that negative relationship interactions are associated with higher alcohol consumption and an increased desire to drink.

2.3 (1 points): The researchers hypothesize that the relation between drinking and negative relationship interactions should not be evident for individuals with high trait self-esteem. Conduct an analysis to address this hypothesis.