Statistical Methods for Discrete Response, Time Series, and Panel Data (W271): Lab 2

Harry Xu and Rich Ung

Strategic Placement of Products in Grocery Stores

Answer **Question 12 of chapter 3 (on page 189 and 190)** of Bilder and Loughin's *"Analysis of Categorical Data with R"*. Here is the background of this analysis, taken as an excerpt from this question:

```
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.
```

Since this question has part a to h, please write down each of the questions in your report so that I can easily follow your answer.

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. Using these data, complete the following:

```r
cereal <- read.csv('cereal_dillons.csv', header=TRUE)
cereal
```

```
##    ID Shelf                               Cereal size_g sugar_g
## 1   1     1     Kellog's Razzle Dazzle Rice Crispies     28      10
## 2   2     1            Post Toasties Corn Flakes     28       2
## 3   3     1                Kellogg's Corn Flakes     28       2
## 4   4     1               Food Club Toasted Oats     32       2
## 5   5     1                     Frosted Cheerios     30      13
## 6   6     1             Food Club Frosted Flakes     31      11
## 7   7     1                          Capn Crunch     27      12
## 8   8     1      Capn Crunch's Peanut Butter Crunch     27       9
## 9   9     1                      Post Honeycomb      29      11
## 10 10     1                Food Club Crispy Rice     33       2
## 11 11     2                 Rice Crispies Treats     30       9
## 12 12     2                     Kellogg's Smacks     27      15
## 13 13     2                Kellogg's Froot Loops     32      15
## 14 14     2      Capn Crunch's Peanut Butter Crunch     27       9
## 15 15     2                     Cinnamon Grahams     30      11
## 16 16     2        Marshmallow Blasted Froot Loops     30      16
```

```
## 17 17     2                                 Koala Coco Krunch     30       13
## 18 18     2                          Food Club Toasted Oats     33       10
## 19 19     2                                   Cocoa Pebbles     29       13
## 20 20     2                                        Oreo O's     27       11
## 21 21     3                           Food Club Raisin Bran     54       17
## 22 22     3                      Post Honey Bunches of Oats     30        6
## 23 23     3                                       Rice Chex     31        2
## 24 24     3                              Kellogg's Corn Pops     31       14
## 25 25     3 Post Morning Traditions - Raisin, Date, Pecan      54       14
## 26 26     3                  Post Shredded Wheat Spoon Size     49        0
## 27 27     3                                         Basic 4     55       14
## 28 28     3                              French Toast Crunch     30       12
## 29 29     3                                Post Raisin Bran     59       20
## 30 30     3                 Food Club Frosted Shreded Wheat     50        1
## 31 31     4                               Total Raisin Bran     55       19
## 32 32     4                          Food Club Wheat Crunch     60        6
## 33 33     4                            Oatmeal Crisp Raisin     55       19
## 34 34     4                           Food Club Bran Flakes     31        5
## 35 35     4                                    Cookie Crisp     30       12
## 36 36     4                    Kellogg's All Bran Original      31        6
## 37 37     4                       Food Club Low Fat Granola     55       14
## 38 38     4                    Oatmeal Crisp Apple Cinnamon     55       19
## 39 39     4 Post Fruit and Fiber - Dates, Raisons, Walnuts     55       17
## 40 40     4                               Total Corn Flakes     30        3
##    fat_g sodium_mg
## 1    0.0       170
## 2    0.0       270
## 3    0.0       300
## 4    2.0       280
## 5    1.0       210
## 6    0.0       180
## 7    1.5       200
## 8    2.5       200
## 9    0.5       220
## 10   0.0       330
## 11   1.5       190
## 12   0.5        50
## 13   1.0       150
## 14   2.5       200
## 15   1.0       230
## 16   0.5       105
## 17   1.0       170
## 18   1.5       150
## 19   1.0       160
## 20   2.5       150
## 21   1.0       280
## 22   1.5       190
## 23   0.0       290
```

```
## 24   0.0        120
## 25   5.0        160
## 26   0.5          0
## 27   3.0        320
## 28   1.0        180
## 29   1.0        300
## 30   1.0          0
## 31   1.0        240
## 32   0.0        300
## 33   2.0        220
## 34   0.5        220
## 35   1.0        180
## 36   1.0         65
## 37   3.0        100
## 38   2.0        260
## 39   3.0        280
## 40   0.0        200
```

a

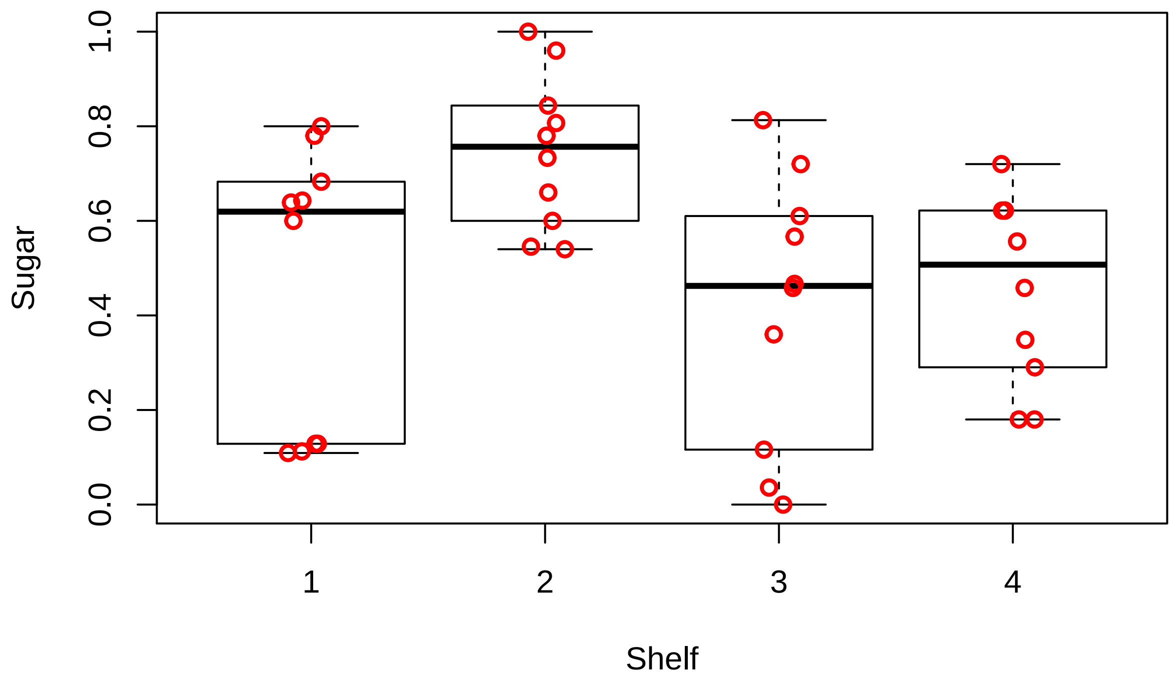
The explanatory variables need to be re-formatted before proceeding further. First, divide each explanatory variable by its serving size to account for the different serving sizes among the cereals. Second, re-scale each variable to be within 0 and 1.[12] Below is code we use to re-format the data after the data file is read into an object named `cereal`:

```
stand01 <- function(x) { (x - min(x))/(max(x) - min(x)) }
cereal2 <- data.frame(Shelf = cereal$Shelf, sugar =
                        stand01(x = cereal$sugar_g/cereal$size_g), fat =
                        stand01(x = cereal$fat_g/cereal$size_g), sodium =
                        stand01(x = cereal$sodium_mg/cereal$size_g))
```

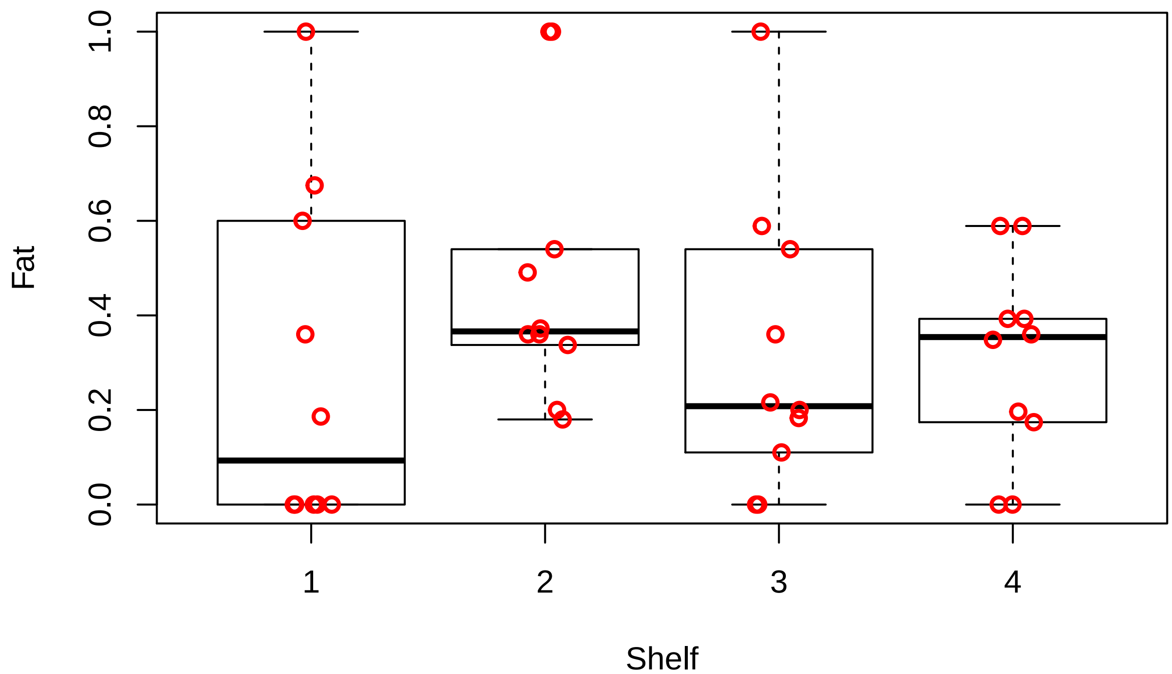
b

Construct side-by-side box plots with dot plots overlaid for each of the explanatory variables. Below is code that can be used for plots involving sugar:

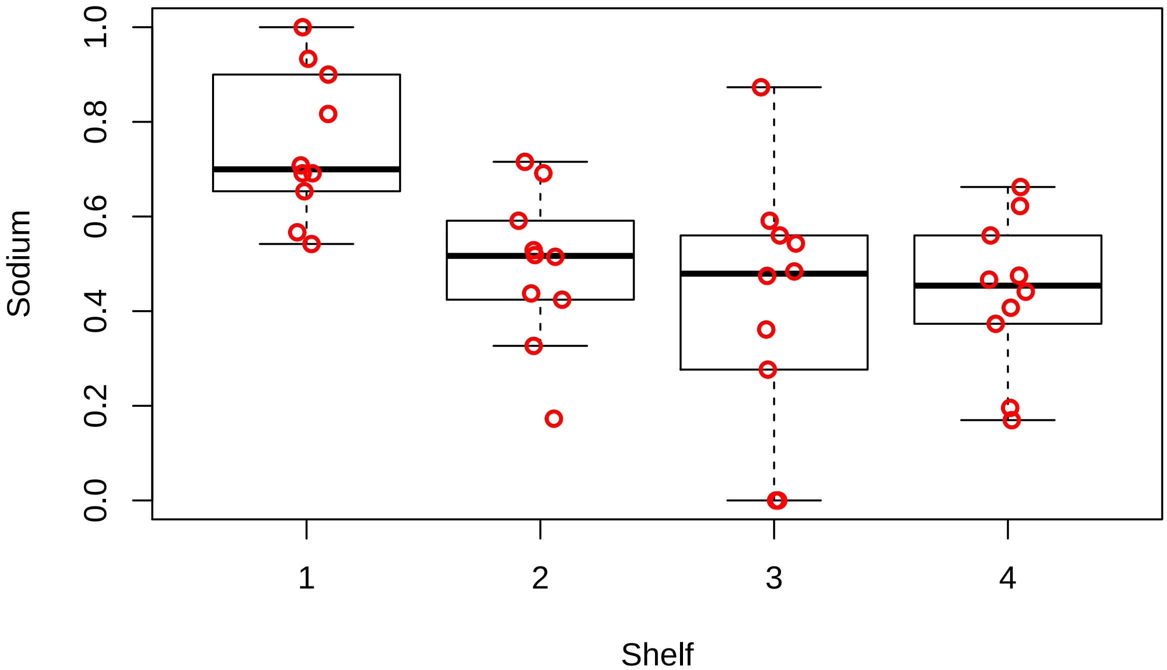
```
boxplot(formula = sugar ~ Shelf, data = cereal2, ylab = "Sugar", xlab = "Shelf", pars = list(ou
stripchart(x = cereal2$sugar ~ cereal2$Shelf, lwd = 2, col = "red", method = "jitter", vertical
```




```
boxplot(formula = fat ~ Shelf, data = cereal2, ylab = "Fat", xlab = "Shelf", pars = list(outpc
stripchart(x = cereal2$fat ~ cereal2$Shelf, lwd = 2, col = "red", method = "jitter", vertical =
```




```
boxplot(formula = sodium ~ Shelf, data = cereal2, ylab = "Sodium", xlab = "Shelf", pars = list
stripchart(x = cereal2$sodium ~ cereal2$Shelf, lwd = 2, col = "red", method = "jitter", vertica
```



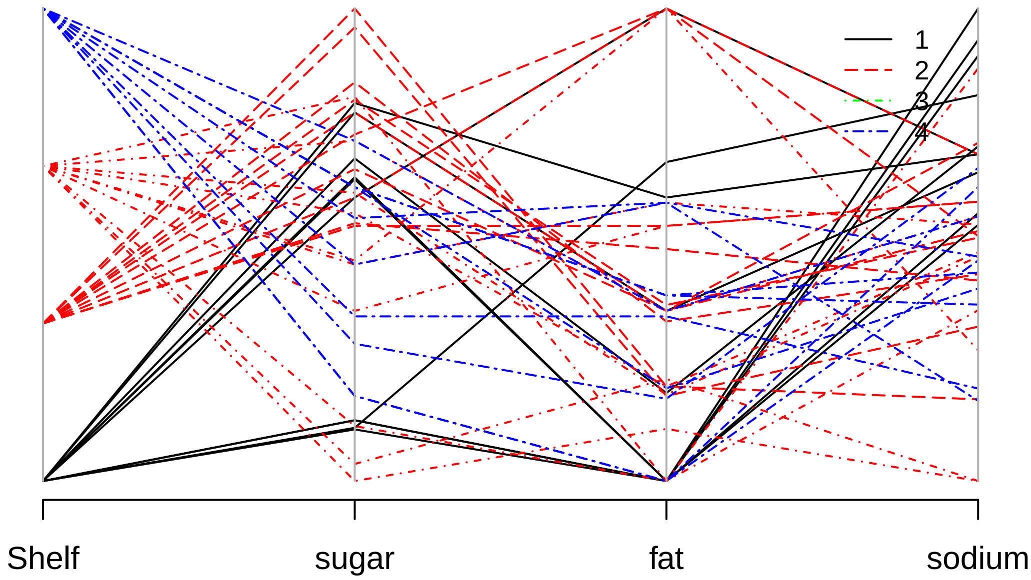
Also, construct a parallel coordinates plot for the explanatory variables and the shelf number. Discuss if possible content differences exist among the shelves.

```
library(package = MASS)

cereal2.colors<-ifelse(test = cereal2$Shelf==1, yes = "black",
                    no = ifelse(test = cereal2$Shelf==2, yes = "red",
                          ifelse(test = cereal2$Shelf==3, yes = "red", no = "blue")))

cereal2.lty<-ifelse(test = cereal2$Shelf==1, yes = "solid",
                    no = ifelse(test = cereal2$Shelf==2, yes = "longdash",
                          ifelse(test = cereal2$Shelf==3, yes = "dotdash", no = "twodash"

parcoord(x = cereal2, col = cereal2.colors, lty = cereal2.lty)  # Plot
legend(x = 3.5, y = 1, legend = c("1", "2", "3", "4"), lty = c("solid", "longdash", "dotdash",
      col=c("black", "red", "green", "blue"), cex=0.8, bty="n")
```



c

The response has values of 1, 2, 3, and 4. Under what setting would it be desirable to take into account ordinality. Do you think this occurs here?

It would be desirable to take into account ordinality when the variable has a natural ordering to their levels. In other words, if response levels can be arranged so that category 1 < category 2 < $\cdots$ < category J in some conceptual scale of measurement (e.g., amount of agreement). Since the shelf has a natural ordering to their levels, bottom (1) to top (4), it would make sense to take into account ordinality.

d

Estimate a multinomial regression model with linear forms of the sugar, fat, and sodium variables. Perform LRTs to examine the importance of each explanatory variable.

```
library(package = MASS)
mod.fit.ord <- polr(formula = as.factor(Shelf) ~ sugar + fat + sodium, data = cereal2, method =
summary(mod.fit.ord)
```

```
##
## Re-fitting to get Hessian

## Call:
## polr(formula = as.factor(Shelf) ~ sugar + fat + sodium, data = cereal2,
##     method = "logistic")
##
## Coefficients:
##           Value Std. Error  t value
## sugar  -1.61101     1.2830 -1.25565
```

```
## fat    -0.05123     0.9657 -0.05305
## sodium -4.85950     1.6302 -2.98094
##
## Intercepts:
##     Value   Std. Error t value
## 1|2 -4.7534  1.4837    -3.2037
## 2|3 -3.3435  1.3810    -2.4210
## 3|4 -1.9823  1.2867    -1.5407
##
## Residual Deviance: 98.52912
## AIC: 110.5291
```

```r
library(package = car)
Anova(mod.fit.ord)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: as.factor(Shelf)
##        LR Chisq Df Pr(>Chisq)
## sugar    1.6794  1  0.1950069
## fat      0.0028  1  0.9577007
## sodium  11.5685  1  0.0006708 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

e

Show that there are no significant interactions among the explanatory variables (including an interaction among all three variables).

```r
library(package = car)
mod.fit.ord2 <- polr(formula = as.factor(Shelf) ~ sugar + fat + sodium
                     + sugar:fat + sugar:sodium + fat:sodium
                     + sugar:fat:sodium
                     , data = cereal2, method = "logistic")
Anova(mod.fit.ord2)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: as.factor(Shelf)
##                  LR Chisq Df Pr(>Chisq)
## sugar              1.1760  1  0.2781685
## fat                0.0419  1  0.8377311
## sodium            11.1699  1  0.0008314 ***
## sugar:fat          0.1014  1  0.7501457
## sugar:sodium       0.3945  1  0.5299556
## fat:sodium         0.2607  1  0.6096643
## sugar:fat:sodium   0.1077  1  0.7427907
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

f

Kellogg's Apple Jacks (http://www.applejacks.com) is a cereal marketed to- ward 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.

```r
summary(mod.fit.ord)
```

```
##
## Re-fitting to get Hessian

## Call:
## polr(formula = as.factor(Shelf) ~ sugar + fat + sodium, data = cereal2,
##     method = "logistic")
##
## Coefficients:
##           Value Std. Error  t value
## sugar  -1.61101     1.2830 -1.25565
## fat    -0.05123     0.9657 -0.05305
## sodium -4.85950     1.6302 -2.98094
##
## Intercepts:
##     Value   Std. Error t value
## 1|2 -4.7534  1.4837    -3.2037
## 2|3 -3.3435  1.3810    -2.4210
## 3|4 -1.9823  1.2867    -1.5407
##
## Residual Deviance: 98.52912
## AIC: 110.5291
```

```r
predict(object = mod.fit.ord, newdata = data.frame(sugar = (((12/28) - min(cereal$sugar_g))/(ma
                                                   fat = (((0.5/28) - min(cereal$fat_g))/(max(
                                                   sodium = (((130/28) - min(cereal$sodium_mg)
```

```
##           1           2           3           4
## 0.009468087 0.028204574 0.094803728 0.867523610
```

g

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.

h

Estimate odds ratios and calculate corresponding confidence intervals for each explanatory variable. Relate your interpretations back to the plots constructed for this exercise.