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

Instructions:

- **Due Date: Monday June 17 4p.m. Pacific Time**
- **Page limit of the pdf report: 20 (not include title and the table of content page**
- Use the margin, linespace, and font size specification below:
  - fontsize=11pt
  - margin=1in
  - line_spacing=single
- Submission:
  - Each group makes one submission to Github; please have one of your team members made the submission
  - Submit 2 files:
    1. A pdf file including the details of your analysis and all the R codes used to produce the analysis. Please do not suppress the codes in your pdf file.
    2. R markdown file used to produce the pdf file
  - Use the following file-naming convensation; fail to do so will receive 10% reduction in the grade:
    - FirstNameLastName1_FirstNameLastName2_FirstNameLastName3_LabNumber.fileExtension
    - For example, if you have three students in the group for Lab Z, and their names are Gerard Kelley, Steve Yang, and Jeffrey Yau, then you should name your file the following
      - GerardKelly_SteveYang_JeffreyYau_LabZ.Rmd
      - GerardKelly_SteveYang_JeffreyYau_LabZ.pdf
  - Although it sounds obvious, please write the name of each members of your group on page 1 of your pdf and Rmd files.
  - This lab can be completed in a group of up to 3 students in your session. Students are encouraged to work in a group for the lab.
- Other general guidelines:
  - If you use libraries and functions for statistical modeling that we have not covered in this course, you have to provide (1) explanation of why such libraries and functions are used instead and (2) reference to the library documentation. Lacking the explanation and reference to the documentation will result in a score of zero for the corresponding question.
  - In this particular lab, simply answer the following questions stated in Question 12 of chapter 3 (on page 189 and 190) of Bilder and Loughin's *"Analysis of Categorical Data with R"*
  - **No need to include introduction, data examination, EDA, and conclusion sections.**
  - Since this question has **part a to h**, please write down each of the questions in your report so that we can easily follow your answers.
- Students are expected to act with regards to UC Berkeley Academic Integrity.

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.

```r
# Load Libraries
library(car)
```

```
## Loading required package: carData
```

```r
library(Hmisc)
```

```
## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

## Registered S3 methods overwritten by 'ggplot2':
##   method         from
##   [.quosures     rlang
##   c.quosures     rlang
##   print.quosures rlang

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:Hmisc':
##
##     src, summarize

## The following object is masked from 'package:car':
##
```

```
##     recode

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(skimr)
```

```
##
## Attaching package: 'skimr'

## The following object is masked from 'package:knitr':
##
##     kable

## The following object is masked from 'package:stats':
##
##     filter
```

```r
library(ggplot2)
library(stargazer)
```

```
##
## Please cite as:

##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```r
library(gmodels) # For cross tabulation (SAS and SPSS style)
library(MASS)
```

```
##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select
```

```r
library(mcprofile)
library(vcd)
```

```
## Loading required package: grid
```

```r
library(nnet)
```

a. The explanatory variables need to be reformatted before proceeding further. - 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.
```

```r
cereal <- read.csv("/Users/eugenetang/Desktop/EugeneTang/Grad School/Berkeley/W271_TimeSeries/

summary(cereal)
```

```
##        ID            Shelf                                       Cereal
##  Min.   : 1.00   Min.   :1.00   Capn Crunch's Peanut Butter Crunch: 2
##  1st Qu.:10.75   1st Qu.:1.75   Food Club Toasted Oats            : 2
##  Median :20.50   Median :2.50   Basic 4                           : 1
##  Mean   :20.50   Mean   :2.50   Capn Crunch                       : 1
##  3rd Qu.:30.25   3rd Qu.:3.25   Cinnamon Grahams                  : 1
##  Max.   :40.00   Max.   :4.00   Cocoa Pebbles                     : 1
##                                 (Other)                           :32
##      size_g         sugar_g          fat_g          sodium_mg
##  Min.   :27.00   Min.   : 0.0   Min.   :0.000   Min.   :  0.0
##  1st Qu.:29.75   1st Qu.: 6.0   1st Qu.:0.500   1st Qu.:157.5
##  Median :31.00   Median :11.0   Median :1.000   Median :200.0
##  Mean   :37.20   Mean   :10.4   Mean   :1.200   Mean   :195.5
##  3rd Qu.:51.00   3rd Qu.:14.0   3rd Qu.:1.625   3rd Qu.:262.5
##  Max.   :60.00   Max.   :20.0   Max.   :5.000   Max.   :330.0
##
```

```r
head (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                 Kellogg'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
```

Two cereals ("Capn Crunch's Peanut Butter Crunch" and "Food Club Toasted Oats") show up twice in the dataset. This is because both cereals appear on two shelves (shelves one and two).

```r
# function to scale variables linearly to be between 0 and 1
rescale01 <- function (x) {(x - min (x)) / (max (x) - min (x))}

# sugar, fat, and sodium contain the transformed variables
cereal2 <- data.frame (Shelf = cereal$Shelf, size_g = cereal$size_g,
                       sugar_g =cereal$sugar_g, fat_g = cereal$fat_g,
                       sodium_mg = cereal$sodium_mg,
                       sugar = rescale01(x = cereal$sugar_g/cereal$size_g),
```

```r
                        fat = rescale01(x = cereal$fat_g/cereal$size_g),
                        sodium = rescale01(x = cereal$sodium_mg/cereal$size_g))

summary(cereal2)
```

```
##      Shelf          size_g         sugar_g          fat_g      
##  Min.   :1.00   Min.   :27.00   Min.   : 0.0   Min.   :0.000  
##  1st Qu.:1.75   1st Qu.:29.75   1st Qu.: 6.0   1st Qu.:0.500  
##  Median :2.50   Median :31.00   Median :11.0   Median :1.000  
##  Mean   :2.50   Mean   :37.20   Mean   :10.4   Mean   :1.200  
##  3rd Qu.:3.25   3rd Qu.:51.00   3rd Qu.:14.0   3rd Qu.:1.625  
##  Max.   :4.00   Max.   :60.00   Max.   :20.0   Max.   :5.000  
##    sodium_mg         sugar              fat              sodium      
##  Min.   :  0.0   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000  
##  1st Qu.:157.5   1st Qu.:0.3339   1st Qu.:0.1582   1st Qu.:0.4200  
##  Median :200.0   Median :0.6000   Median :0.3542   Median :0.5354  
##  Mean   :195.5   Mean   :0.5209   Mean   :0.3476   Mean   :0.5240  
##  3rd Qu.:262.5   3rd Qu.:0.7200   3rd Qu.:0.5400   3rd Qu.:0.6696  
##  Max.   :330.0   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000  
```

```r
head(cereal2)
```

```
##   Shelf size_g sugar_g fat_g sodium_mg     sugar   fat    sodium
## 1     1     28      10     0       170 0.6428571 0.000 0.5666667
## 2     1     28       2     0       270 0.1285714 0.000 0.9000000
## 3     1     28       2     0       300 0.1285714 0.000 1.0000000
## 4     1     32       2     2       280 0.1125000 0.675 0.8166667
## 5     1     30      13     1       210 0.7800000 0.360 0.6533333
## 6     1     31      11     0       180 0.6387097 0.000 0.5419355
```

The sugar, fat, and sodium columns are all now between 0 and 1 and have been scaled to account for the cereal box size.

b. Construct side-by-side box plots with dot plots overlaid for each of the explanatory variables.

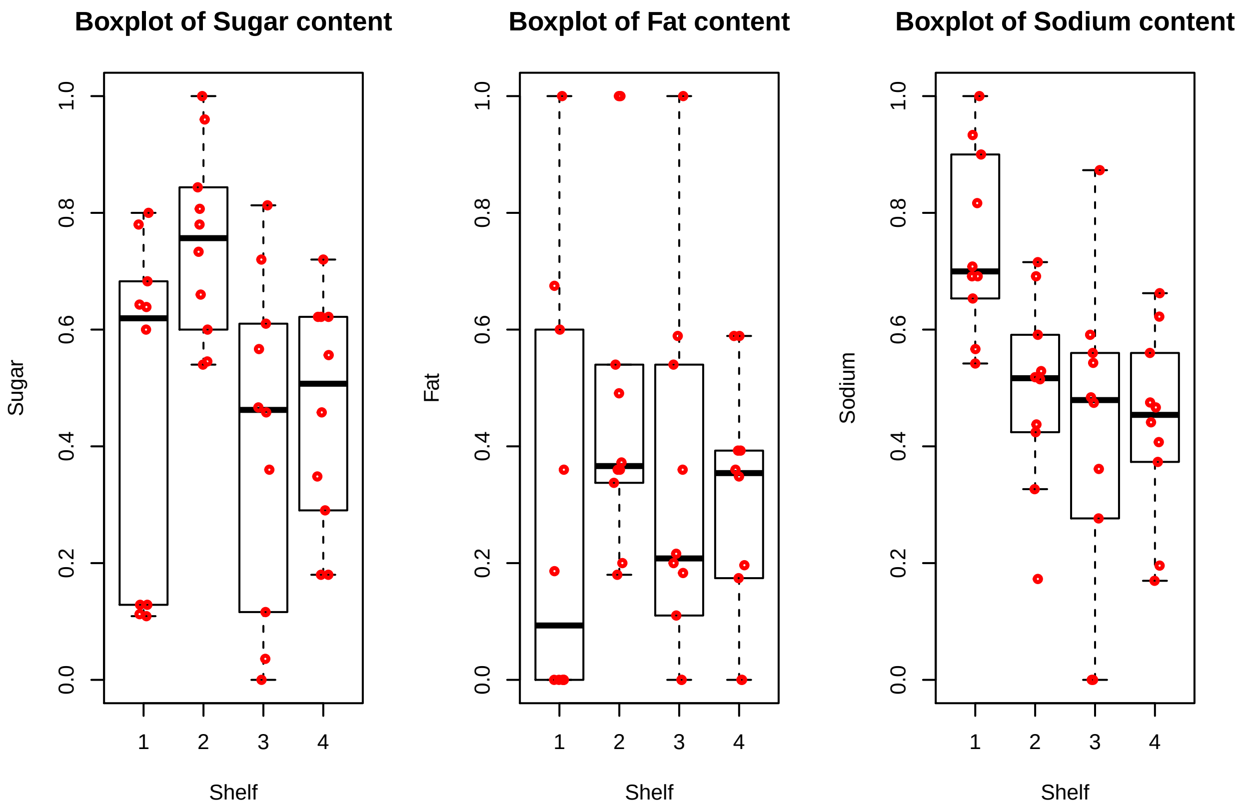
```
- Also, construct a **parallel coordinates plot** for the explanatory variables and the shelf r
```

```r
par(mfrow=c(1,3))
boxplot (sugar ~ Shelf, data = cereal2, ylab = " Sugar ",
         xlab = " Shelf ", outpch = NA,
         main = "Boxplot of Sugar content")
stripchart (sugar ~ Shelf, lwd = 2, col = " red ", vertical = TRUE,
            pch = 1, add = TRUE, ylab = " Sugar ", xlab = " Shelf ",
            data=cereal2, method ="jitter")

boxplot (fat ~ Shelf, data = cereal2, ylab = " Fat ",
         xlab = " Shelf ", outpch = NA,
         main = "Boxplot of Fat content")
stripchart (fat ~ Shelf, lwd = 2, col = " red ", vertical = TRUE,
```

```
            pch = 1, add = TRUE, ylab = " Fat ", xlab = " Shelf ",
            data=cereal2, method ="jitter")

boxplot (sodium ~ Shelf, data = cereal2, ylab = " Sodium ",
         xlab = " Shelf ", outpch = NA,
         main = "Boxplot of Sodium content")
stripchart (sodium ~ Shelf, lwd = 2, col = " red ", vertical = TRUE,
            pch = 1, add = TRUE, ylab = " Sodium  ", xlab = " Shelf ",
            data=cereal2, method ="jitter")
```

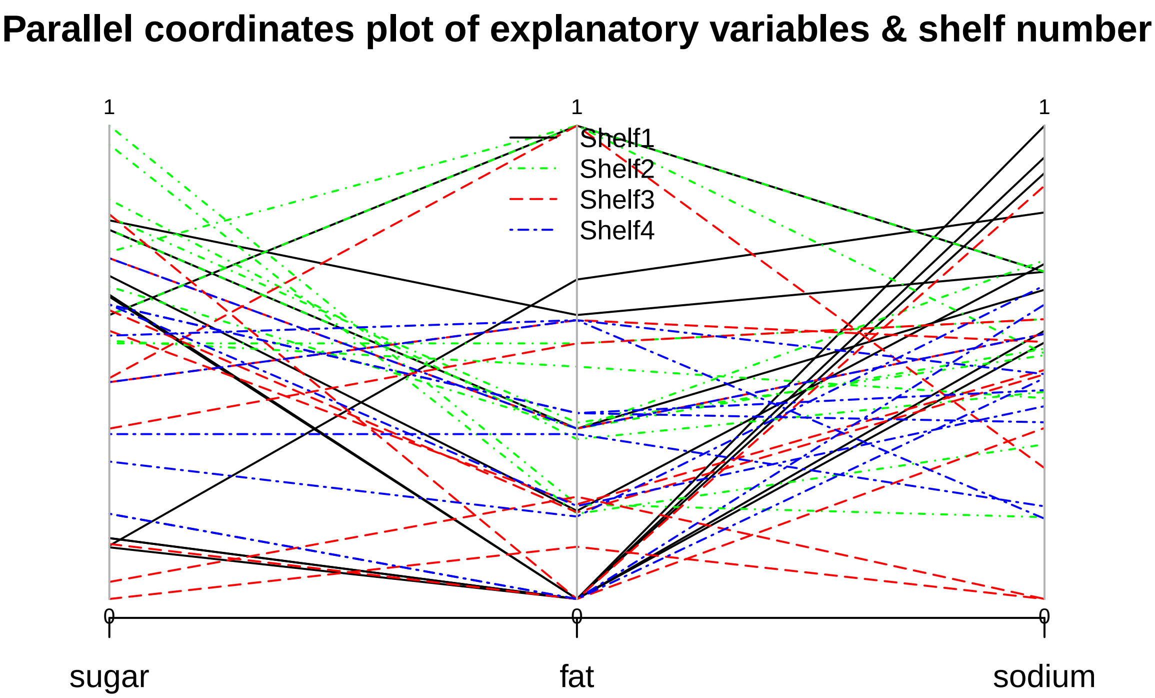



```
# Colors by condition:
df2 <- data.frame (sugar = cereal2$sugar, fat = cereal2$fat, sodium = cereal2$sodium)

# Set cereal colors by shelf number
cereal.colors<-ifelse(test = cereal2$Shelf=="1", yes = "black",
                no = ifelse(test =cereal2$Shelf=="2", yes = "green",
                            no = ifelse(test =cereal2$Shelf=="3", yes = "red",
                                        no = "blue")))
# Set line type by shelf number
cereal.lty<-ifelse(test =  cereal2$Shelf=="1", yes = "solid",
                no = ifelse(test =  cereal2$Shelf=="2", yes = "dotdash",
                            no = ifelse(test =  cereal2$Shelf=="3", yes = "longdash",
                                        no = "twodash")))

# Use parcoord to plot
```

```r
parcoord(x = df2, col = cereal.colors, lty = cereal.lty,
         var.label=TRUE,
         main="Parallel coordinates plot of explanatory variables & shelf number")
legend("top", legend = c("Shelf1", "Shelf2", "Shelf3", "Shelf4"),
       lty=c("solid","dotdash","longdash","twodash"),
col=c("black","green", "red","blue"),cex=0.8, bty="n")
```



From the boxplots and parallel coordinates plots, there seem to be content differences among the shelves. From the boxplots and parallel coordinate plots, we can see that shelf 2 generally contains higher sugar cereals while shelf 1 seems to contain higher sodium cereals (minimum sodium for shelf 1 is over 0.5). Shelf 4 seems to have average-to-low sugar, fat, and sodium levels as none of them exceed 0.8. Shelf 3 seems to have a large range in the sugar, fat, and sodium contents.

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 that this setting occurs here?

The responses 1,2,3,4 here are categorical (shelf 1, 2, 3, 4). When there is a natural ranking to the response, then it is desirable to consider ordinality. For example, when the shelf a cereal is put on depends on the ordering by some metric (e.g. shelves 1-4 corresponds to excellent, good, medium, and bad quality (or price, or popularity)), then we should account for ordinality. We do not have enough information about the shelves (e.g. which shelf is the most "popular", if shelf 1 has the most popular cereals and shelf 4 has the least popular cereals), so we should not assume ordinality here. There is no ranking to the shelves, so they shouldn't be considered ordinal here.

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.

```r
# run regression
mod.nominal1 <- multinom(formula = Shelf ~ sugar + fat + sodium, data = cereal2)
```

```
## # weights:  20 (12 variable)
## initial  value 55.451774
## iter  10 value 37.329384
## iter  20 value 33.775257
## iter  30 value 33.608495
## iter  40 value 33.596631
## iter  50 value 33.595909
## iter  60 value 33.595564
## iter  70 value 33.595277
## iter  80 value 33.595147
## final  value 33.595139
## converged
```

```r
summary(mod.nominal1)
```

```
## Call:
## multinom(formula = Shelf ~ sugar + fat + sodium, data = cereal2)
##
## Coefficients:
##   (Intercept)      sugar        fat    sodium
## 2    6.900708   2.693071  4.0647092 -17.49373
## 3   21.680680 -12.216442 -0.5571273 -24.97850
## 4   21.288343 -11.393710 -0.8701180 -24.67385
##
## Std. Errors:
##   (Intercept)    sugar      fat   sodium
## 2    6.487408 5.051689 2.307250 7.097098
## 3    7.450885 4.887954 2.414963 8.080261
## 4    7.435125 4.871338 2.405710 8.062295
##
## Residual Deviance: 67.19028
## AIC: 91.19028
```

```r
# Run LRT
Anova(mod.nominal1)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Shelf
##        LR Chisq Df Pr(>Chisq)
## sugar   22.7648  3  4.521e-05 ***
## fat      5.2836  3     0.1522
## sodium  26.6197  3  7.073e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In the regression, the coefficients of sugar, fat, and sodium are all negative for shelves 3 and 4. The coefficients are positive for sugar and fat for shelf 2, but is negative for sodium for shelf 2. The LRT test is performed by the Anova function. From the results of the LRT test, Sugar and sodium are both statistically significant at a 0.001 level. Sugar has a p-value of 0.00004521 and sodium has a p-value of 0.000007073. Fat, on the other hand, is not stastically significant even at the 0.1 level.

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

```
mod.nominal2 <- multinom(formula = Shelf ~ sugar + fat + sodium +
                           sugar:fat + sugar:sodium + fat:sodium +
                           sugar:fat:sodium, data = cereal2)
```

```
## # weights:  36 (24 variable)
## initial  value 55.451774
## iter  10 value 36.170336
## iter  20 value 31.166546
## iter  30 value 29.963705
## iter  40 value 28.414027
## iter  50 value 27.891712
## iter  60 value 27.763967
## iter  70 value 27.622579
## iter  80 value 27.438263
## iter  90 value 27.015534
## iter 100 value 26.772481
## final  value 26.772481
## stopped after 100 iterations
```

```
summary(mod.nominal2)
```

```
## Call:
## multinom(formula = Shelf ~ sugar + fat + sodium + sugar:fat +
##     sugar:sodium + fat:sodium + sugar:fat:sodium, data = cereal2)
##
## Coefficients:
##   (Intercept)      sugar       fat     sodium sugar:fat sugar:sodium
## 2   -4.563627   8.944868 22.063003   1.030077  35.60873   -12.250084
## 3   24.498320 -22.248456 35.981865 -27.899087 -17.12487    13.253103
## 4   27.246742 -21.852777  7.298799 -29.106797  41.08251     2.887805
##   fat:sodium sugar:fat:sodium
## 2  -23.75955        -55.88455
## 3  -59.54150         37.71571
## 4  -30.85250        -22.59552
##
## Std. Errors:
##   (Intercept)    sugar       fat   sodium sugar:fat sugar:sodium
## 2    25.21113 29.72894  96.57821 27.29915  135.1117     31.98647
## 3    22.83750 25.81043 101.17670 24.61166  150.1228     26.89827
## 4    22.80359 26.00692 100.83444 24.51538  150.6750     28.86631
##   fat:sodium sugar:fat:sodium
## 2   116.0776         158.8091
## 3   138.0237         212.2222
## 4   138.5448         217.3953
##
```

```
## Residual Deviance: 53.54496
## AIC: 101.545
```

```
Anova(mod.nominal2)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Shelf
##                 LR Chisq Df Pr(>Chisq)
## sugar            19.2525  3  0.0002424 ***
## fat               6.1167  3  0.1060686
## sodium           30.8407  3  9.183e-07 ***
## sugar:fat         3.2309  3  0.3573733
## sugar:sodium      3.0185  3  0.3887844
## fat:sodium        3.1586  3  0.3678151
## sugar:fat:sodium  2.5884  3  0.4595299
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on an LRT test, we find that there are no significant interactions among the explantory variables at a 0.1 level (all p-values are under above 0.1 in the LRT test).

f. 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.

```
# function to allow for rescaling in the same way the data was originally scaled
rescale01_v2 <- function (x, values) {(x - min (values)) / (max (values) - min (values))}

apple_jacks_size = 28
apple_jacks_sugar = 12
apple_jacks_fat = 0.5
apple_jacks_sodium = 130

# add Kellogg's Apple Jacks into data frame
apple_jacks <- data.frame(sugar = rescale01_v2(apple_jacks_sugar/apple_jacks_size, cereal$sugai
                     fat = rescale01_v2(apple_jacks_fat/apple_jacks_size, cereal$fat_g/cereal$s
                     sodium = rescale01_v2(apple_jacks_sodium/apple_jacks_size, cereal$sodium_r

# predict the shelf probabilities for Apple Jacks
predict(object = mod.nominal1, newdata = apple_jacks, type = "probs" )
```

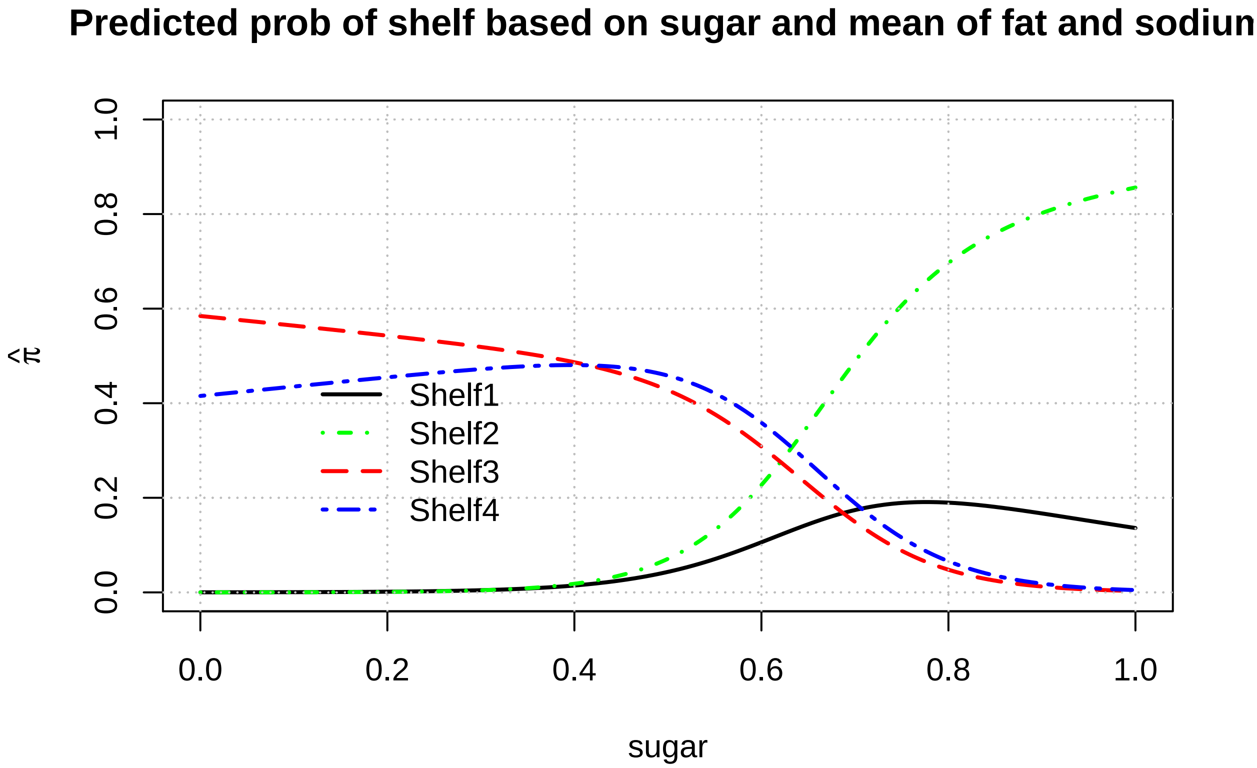
```
##          1          2          3          4
## 0.05326849 0.47194264 0.20042742 0.27436145
```

The shelf probabilities are displayed above. Based on the information of Apple Jacks, it is most likely to be put on shelf 2 with probability of 0.4719.

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.

```
curve(expr = predict(object = mod.nominal1, newdata = data.frame(
  sugar = x, fat = mean(cereal2$fat), sodium = mean(cereal2$sodium)),
  type = "probs")[,1], ylab = expression(hat(pi)), ylim = c(0,1), xlab = "sugar",
  xlim = c(min(cereal2$sugar), max(cereal2$sugar)), col = "black", lty = "solid",
  lwd = 2, n = 1000, panel.first = grid(col = "gray", lty = "dotted"))
curve(expr = predict(object = mod.nominal1, newdata = data.frame(
  sugar = x, fat = mean(cereal2$fat), sodium = mean(cereal2$sodium)),
  type = "probs")[,2], ylab = expression(hat(pi)),xlab = "sugar",
  xlim = c(min(cereal2$sugar), max(cereal2$sugar)), col = "green", lty = "dotdash",
  lwd = 2, n = 1000, add = TRUE, panel.first = grid(col = "gray", lty = "dotted"))
curve(expr = predict(object = mod.nominal1, newdata = data.frame(
  sugar = x, fat = mean(cereal2$fat), sodium = mean(cereal2$sodium)),
  type = "probs")[,3], ylab = expression(hat(pi)),xlab = "sugar",
  xlim = c(min(cereal2$sugar), max(cereal2$sugar)), col = "red", lty = "longdash", lwd = 2, n =
  add = TRUE, panel.first = grid(col = "gray", lty = "dotted"))
curve(expr = predict(object = mod.nominal1, newdata = data.frame(
  sugar = x, fat = mean(cereal2$fat), sodium = mean(cereal2$sodium)),
  type = "probs")[,4],ylab = expression(hat(pi)), xlab = "sugar",
  xlim = c(min(cereal2$sugar), max(cereal2$sugar)), col = "blue", lty = "twodash", lwd = 2, n =
  add = TRUE, panel.first = grid(col = "gray", lty = "dotted"))
legend(x = 0.1, y = 0.5, legend=c("Shelf1", "Shelf2", "Shelf3", "Shelf4"), lty=c("solid","dotda
title(main ="Predicted prob of shelf based on sugar and mean of fat and sodium")
```



By using the mean of fat per cereal and sodium per cereal, when sugar per cereal increases, the estimated probability of being placed on Shelves 1 and 2 increases. Notably, the estimated probability of the cereal being on shelf 2 is fairly high when the sugar content of a cereal is high (above 0.75). On the other hand, when a cereal has low sugar content, the estimated probability of being placed on shelves 3 and 4 is fairly high.

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

For the odds ratios, we compute the odds ratios for each shelf against the baseline (shelf 1). For the value of c in each comparison, we use 0.2, which is approximately the standard deviation of the sugar, fat, and sodium variables. We show the results for the odds ratios given a positive c and a negative c to better interpret the effect of both increases and decreases in odds.

```
# compute standard deviation of each variable just to get a sense
sqrt(var(cereal2$sugar))
```

```
## [1] 0.2692078
```

```
sqrt(var(cereal2$fat))
```

```
## [1] 0.2990292
```

```
sqrt(var(cereal2$sodium))
```

```
## [1] 0.2298359
```

```
# we will use c = 0.2 as an approximation to c.
c = 0.2

beta.hats = coefficients(mod.nominal1)

# compute odds ratios
exp(c * beta.hats[,"sugar"])
```

```
##          2          3          4
## 1.71363042 0.08687471 0.10241297
```

```
exp(-c * beta.hats[,"sugar"])
```

```
##          2          3          4
##  0.5835564 11.5108300  9.7643885
```

```
exp(c * beta.hats[,"fat"])
```

```
##         2         3         4
## 2.2545307 0.8945581 0.8402771
```

```
exp(-c * beta.hats[,"fat"])
```

```
##         2         3         4
## 0.4435513 1.1178704 1.1900836
```

```r
exp(c * beta.hats[,"sodium"])
```

```
##           2           3           4
## 0.030235290 0.006766989 0.007192119
```

```r
exp(-c * beta.hats[,"sodium"])
```

```
##         2         3         4
##  33.07393 147.77621 139.04107
```

We now interpret the odds as follows:

- The estimated odds of a cereal being on shelf 2 vs. shelf 1 change by 1.71 times for a 0.2 *increase* in the sugar content holding the other variables constant.
- The estimated odds of a cereal being on shelf 3 vs. shelf 1 change by 11.51 times for a 0.2 *decrease* in the sugar content holding the other variables constant.
- The estimated odds of a cereal being on shelf 4 vs. shelf 1 change by 9.76 times for a 0.2 *decrease* in the sugar content holding the other variables constant.
- The estimated odds of a cereal being on shelf 2 vs. shelf 1 change by 2.25 times for a 0.2 *increase* in the fat content holding the other variables constant.
- The estimated odds of a cereal being on shelf 3 vs. shelf 1 change by 1.12 times for a 0.2 *decrease* in the fat content holding the other variables constant.
- The estimated odds of a cereal being on shelf 4 vs. shelf 1 change by 1.19 times for a 0.2 *decrease* in the fat content holding the other variables constant.
- The estimated odds of a cereal being on shelf 2 vs. shelf 1 change by 33.07 times for a 0.2 *decrease* in the sodium content holding the other variables constant.
- The estimated odds of a cereal being on shelf 3 vs. shelf 1 change by 147.76 times for a 0.2 *decrease* in the sodium content holding the other variables constant.
- The estimated odds of a cereal being on shelf 4 vs. shelf 1 change by 139.04 times for a 0.2 *decrease* in the sodium content holding the other variables constant.

Next we compute the corresponding confidence intervals for the odds ratios and provide their interpretations below.

```r
beta_cis = confint(mod.nominal1, level = 0.95)
odds_ratio_cis = exp(c * beta_cis)
odds_ratio_cis
```

```
## , , 2
##
##                   2.5 %     97.5 %
## (Intercept) 0.312586045 50.5598912
## sugar       0.236546235 12.4141871
## fat         0.912576279  5.5698453
## sodium      0.001871984  0.4883444
##
## , , 3
##
##                    2.5 %       97.5 %
## (Intercept) 4.1183097759 1417.7539318
```

```
## sugar       0.0127869380    0.5902285
## fat         0.3471238477    2.3053274
## sodium      0.0002849767    0.1606873
##
## , , 4
##
##                     2.5 %       97.5 %
## (Intercept) 3.8311045190 1302.6868805
## sugar       0.0151724908    0.6912784
## fat         0.3272454530    2.1576023
## sodium      0.0003050206    0.1695839
```

- With 95% confidence, the estimated odds of a cereal being on shelf 2 vs. shelf 1 change by 0.24 to 12.41 times for a 0.2 *increase* in the sugar content holding the other variables constant.
- With 95% confidence, the estimated odds of a cereal being on shelf 3 vs. shelf 1 change by 0.013 to 0.590 times for a 0.2 *increase* in the sugar content holding the other variables constant.
- With 95% confidence, the estimated odds of a cereal being on shelf 4 vs. shelf 1 change by 0.015 to 0.691 times for a 0.2 *increase* in the sugar content holding the other variables constant.
- With 95% confidence, the estimated odds of a cereal being on shelf 2 vs. shelf 1 change by 0.91 to 5.57 times for a 0.2 *increase* in the fat content holding the other variables constant.
- With 95% confidence, the estimated odds of a cereal being on shelf 3 vs. shelf 1 change by 0.35 to 2.31 times for a 0.2 *increase* in the fat content holding the other variables constant.
- With 95% confidence, the estimated odds of a cereal being on shelf 4 vs. shelf 1 change by 0.33 to 2.16 times for a 0.2 *increase* in the fat content holding the other variables constant.
- With 95% confidence, the estimated odds of a cereal being on shelf 2 vs. shelf 1 change by 0.0019 to 0.4883 times for a 0.2 *increase* in the sodium content holding the other variables constant.
- With 95% confidence, the estimated odds of a cereal being on shelf 3 vs. shelf 1 change by 0.0003 to 0.1607 times for a 0.2 *increase* in the sodium content holding the other variables constant.
- With 95% confidence, the estimated odds of a cereal being on shelf 4 vs. shelf 1 change by 0.0003 to 0.1700 times for a 0.2 *increase* in the sodium content holding the other variables constant.

First, we observe that the confidence intervals are generally pretty wide, especially around the fat variable, which crosses the 1 boundary. This matches with the results earlier that fat was not a statistically significant variable at the 0.05 level.

These results match up with the general trends we noticed above. First, we saw in the boxplots and parallel coordinates plot that shelf 1 tends to have higher-sodium cereals, a trend we see in the odds ratios as well. Next, we noticed in the boxplots, parallel coordinate plots, and estimated probability graphs that shelf 2 tends to have higher-sugar cereals, a trend we see in the odds ratios as well. Similarly in the graphs and the odds ratios, we see that shelves 3 and 4 tend to have lower-sugar cereals. There was no clear trend in the fat levels, which we also observe in the odds ratios.