# Lab 1, Short Questions

## Contents



```
#install.packages("GGally")
#install.packages("hrbrthemes")
#install.packages("ggthemes")
library(tidyverse)
```

```
## Warning in system("timedatectl", intern = TRUE): running command 'timedatectl'
## had status 1
```

```
library(patchwork)
library(GGally)
library(viridis)
library(hrbrthemes)
library(gridExtra)
library(ggthemes)

## for multinomial log-linear models.
if(!"nnet"%in%rownames(installed.packages())) {install.packages("nnet")}
library(nnet)

## To use plor()
if(!"MASS"%in%rownames(installed.packages())) {install.packages("MASS")}
library(MASS)

## provide useful functions to facilitate the application and interpretation of regression ana
if(!"car"%in%rownames(installed.packages())) {install.packages("car")}
library(car)
```

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

These questions are taken from Question 12 of chapter 3 of the textbook(Bilder and Loughin's "Analysis of Categorical Data with R.

> *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.*

```
cereal <- read_csv('./cereal_dillons.csv')
```

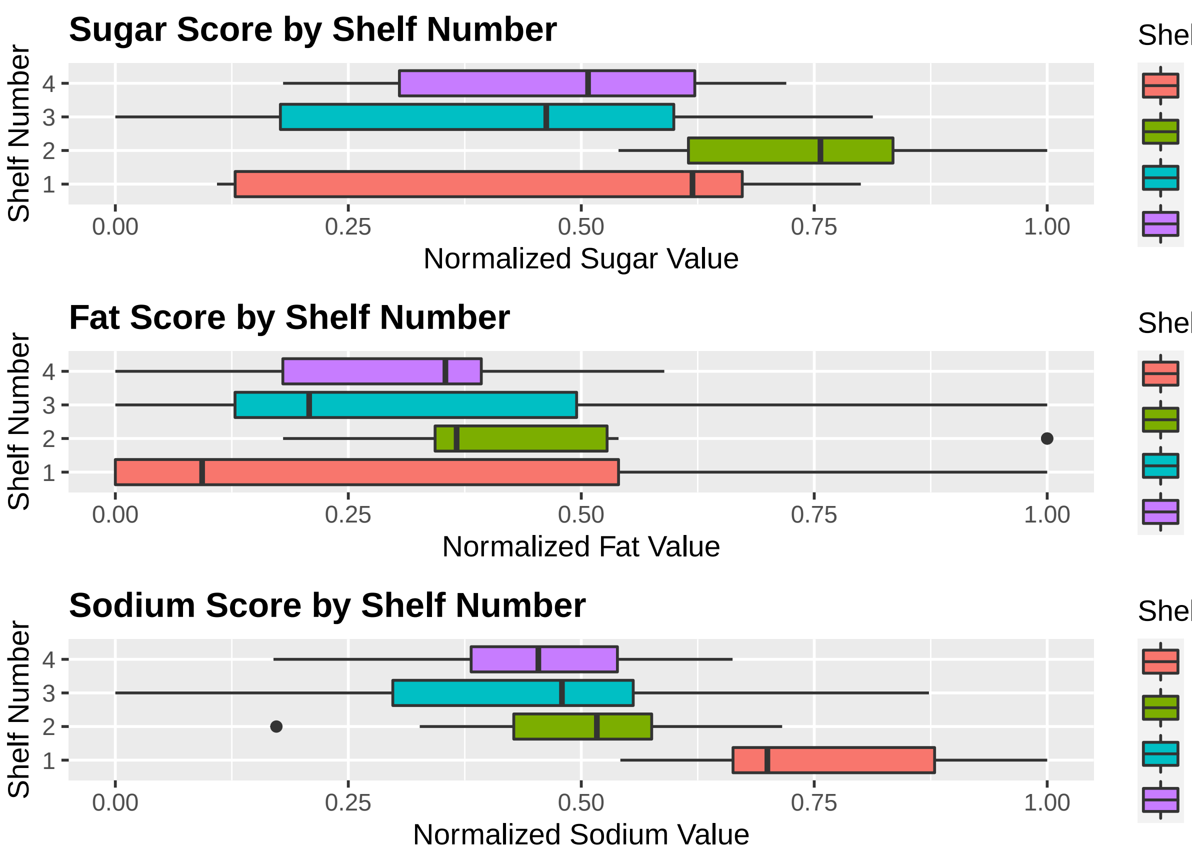
## 1.1 Recode Data

(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.

```
# Set up the function for normalizing data between 0 and 1
stand01 <- function (x) {
  (x - min(x)) / ( max(x) - min(x))
  }

#Establish new dataframe
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 ))


# Setting shelf up as a factor
cereal2 <- cereal2 %>%
  mutate(
    Shelf = factor(Shelf)
      )

#view(cereal2)
```

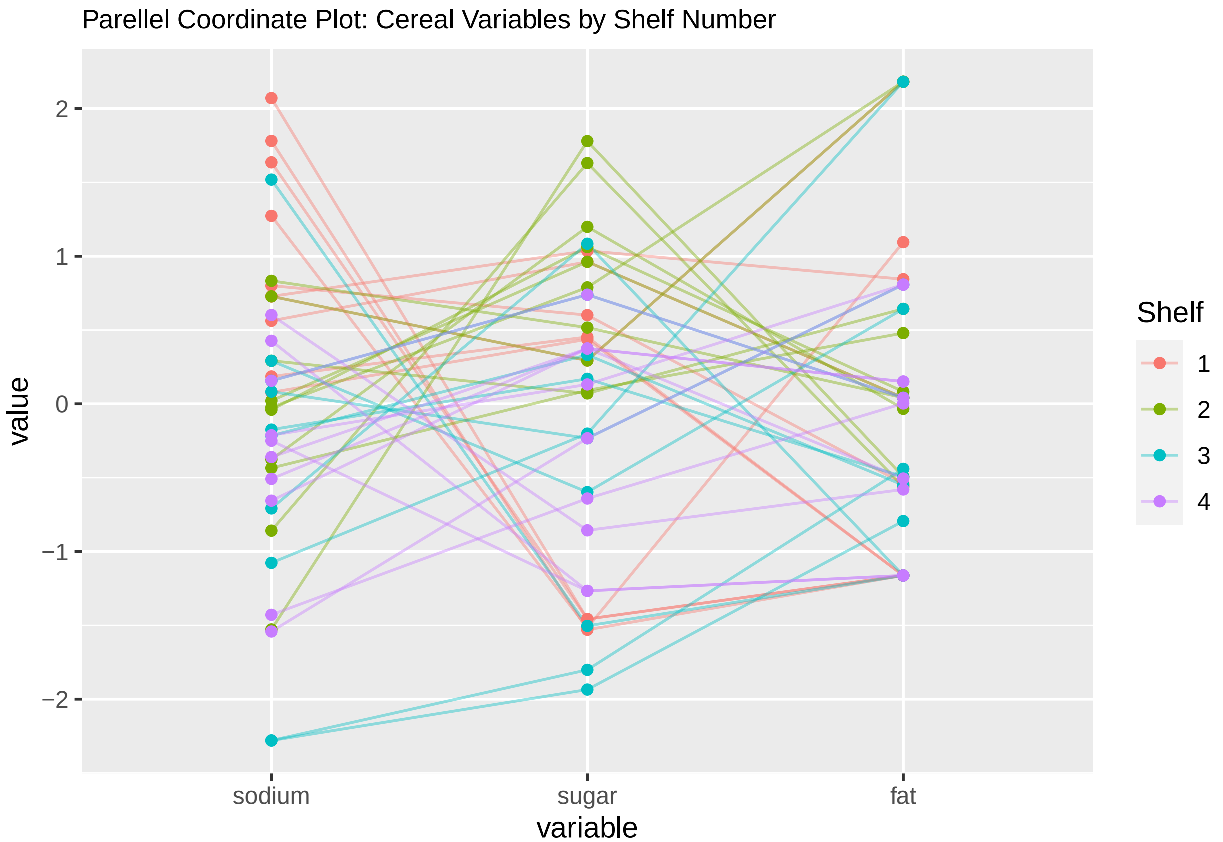
```
# Use box plots to examine response variable with quantitative data
p1 <- cereal2 %>%
  ggplot(aes(Shelf, sugar)) + geom_boxplot(aes(fill = Shelf)) +
  coord_flip() + ggtitle("Sugar Score by Shelf Number") +
```

```
  theme(plot.title = element_text(lineheight = 1, face = "bold")) +
  ylab("Normalized Sugar Value") + xlab("Shelf Number")

p2 <- cereal2 %>%
  ggplot(aes(Shelf, fat)) + geom_boxplot(aes(fill = Shelf)) +
  coord_flip() + ggtitle("Fat Score by Shelf Number") +
  theme(plot.title = element_text(lineheight = 1, face = "bold")) +
  ylab("Normalized Fat Value") + xlab("Shelf Number")

p3 <- cereal2 %>%
  ggplot(aes(Shelf, sodium)) + geom_boxplot(aes(fill = Shelf)) +
  coord_flip() + ggtitle("Sodium Score by Shelf Number") +
  theme(plot.title = element_text(lineheight = 1, face = "bold")) +
  ylab("Normalized Sodium Value") + xlab("Shelf Number")


grid.arrange(p1, p2, p3, nrow = 3, ncol = 1)
```




Shelf number 2 appears to have higher sugar scores than the other shelves, although the sugar values for the other shelves have a wide range of values. Fat scores for many shelves are comparable but there is one outlier on shelf 2 that is worth further investigation. The sodium score is extremely high on shelf 1 and there is an outlier on shelf 2. Both of these factors warrant further investigation.

```r
ggparcoord(cereal2,
    columns = 2:4,
    alphaLines = .4,
    groupColumn = 1, order = "anyClass",
    showPoints = TRUE,
    title = "Parellel Coordinate Plot: Cereal Variables by Shelf Number") +
  theme(
    plot.title = element_text(size = 10)
  )
```




shelf 1 appears to have higher levels of sodium, while shelf 2 appears to have higher levels of sugar. Shelf 3 does have some low values for sodium and sugar but it does not appear that the other values are lower than the other observations. Therefore, there may be content differences between shelf 1 levels of sodium and shelf 2 levels of sugar relative to the other levels.

### 1.2 Evaluate Ordinal vs. Categorical

(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).

The setting under which it would be desirable to take into account ordinality is that situation where

1 represents the lowest physical level of a shelf as well as the lowest preferred level by shoppers. For example, a shelf one would be considered undersirable while a level of 4 would be desirable by all shoppers. Although the dimensions of the shelves matter, it is typically the case the people prefer shelves 2 and 3 because they are easier to access relative to shelves 1 and 4. Under ideal conditions, Dillons would be able to order shelf levels by shopper preference (e.g. level 1 = really not desirable, level 2 = not desirable, level 3 = desirable, level 4 = really desirable).

```r
# Model with normalized explanatory variables
model_cereal_shelves_linear <- multinom(formula = Shelf ~ sodium + sugar + fat,
                                        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(model_cereal_shelves_linear)
```

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

```r
# Model with normalized explanatory variables + interactions terms
model_cereal_shelves_quadratic <- multinom(formula = Shelf ~ sodium + sugar + fat +
                                             sodium:sugar + sodium:fat + sugar:fat,
                                           data = cereal2)
```

```
## # weights:  32 (21 variable)
```

```
## initial  value 55.451774
## iter  10 value 36.108903
## iter  20 value 31.221758
## iter  30 value 30.183143
## iter  40 value 28.725940
## iter  50 value 28.510281
## iter  60 value 28.344031
## iter  70 value 28.286691
## iter  80 value 28.187825
## iter  90 value 28.133526
## iter 100 value 28.066673
## final  value 28.066673
## stopped after 100 iterations
```

```
summary(model_cereal_shelves_quadratic)
```

```
## Call:
## multinom(formula = Shelf ~ sodium + sugar + fat + sodium:sugar +
##     sodium:fat + sugar:fat, data = cereal2)
##
## Coefficients:
##   (Intercept)       sodium     sugar       fat sodium:sugar sodium:fat
## 2   -1.988003   0.03256223  12.63023 25.885108   -25.645040  -41.02996
## 3   26.381935 -30.29171281 -22.80111 20.460284    12.891752  -35.73871
## 4   27.530652 -28.60684017 -17.17032  4.312374    -8.362321  -31.40412
##   sugar:fat
## 2 13.555737
## 3  5.754849
## 4 33.926682
##
## Std. Errors:
##   (Intercept)   sodium    sugar      fat sodium:sugar sodium:fat sugar:fat
## 2    24.17632 25.65306 27.66281 39.72160     29.68042   38.88213  35.44831
## 3    21.62372 23.36754 24.85328 39.71878     27.83209   42.84818  33.43987
## 4    21.52055 23.13444 24.91880 38.89680     28.91377   41.87341  33.68466
##
## Residual Deviance: 56.13335
## AIC: 98.13335
```

For the block of coefficients in the models, the first row compares the coefficients for Shelf=2 to our baseline of Shelf=1. The second row compares the coefficients for Shelf=3 to our our baseline of Shelf=1. Finally, the third row compares the coefficients for Shelf=4 to our baseline of Shelf=1.

When sodium changes by one standard deviation in the first model, the log odds of being on the 2nd versus the 1st shelf decrease by 17.49. Similarly, when sodium changes by one standard deviation, to log odds of being on the 3rd versus the 1st shelf and the 4th versus the first shelf decreased by 25.0 and 24.7, respectively. When sugar changes by one standard deviation, the log odds of being on the 2nd shelf increases by 2.7; the log odds of being on the 3rd shelf decreases by 12.2; and the log odds of being the 4th shelf decreases by 11. Finally, when fat increases by one standard deviation, the log odds of being on the 2nd shelf increases by 4.1; the log odds of being on the 3rd shelf decreases by 0.6; and the log odds of being on the 4 shelf decrease by 0.9.

```r
# LRT for model with normalized explanatory variables only
lrt_cereal_main_effects <- Anova(model_cereal_shelves_linear)
lrt_cereal_main_effects
```

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

```r
# LRT for model with normalized explanatory variables + interaction terms
lrt_cereal_quadratic_effects <- Anova(model_cereal_shelves_quadratic)
lrt_cereal_quadratic_effects
```

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

We performed an LRT using Anova() to explore if a given explanatory variable $x_r$ statistically significant over all response categories in our multinomial model (shelves: 1, 2, 3, 4). The null hypothesis is testing whether all coefficients in the odds ratios for a particular predictor are zero or not.

Our null hypothesis is

$$H_0 : B_{2r} = B_{3r} = 0$$

While our alternative hypothesis is:

$$H_a : B_{2r} \neq 0 or B_{3r} \neq 0$$

In both models the small p-values indicate that sodium and sugar are statistically significant and affect the probability of having the cereal being placed on a shelf level. The p-value for the fat coefficient is quite large, meaning we fail to reject the null hypothesis that the fat coefficient is zero and insufficient evidence indates that fat content is not essential to our response variable.

Similarly, the p-values for all interaction coefficients is quite large, meaning we also fail to reject the null hypothesis that all interactions are zero.

### 1.3 Where do you think Apple Jacks will be placed?

(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.

```
# Make lists out of original dataset
list <- as.list(cereal)
# make lists for every column of interest
size_list <- list$size_g
sugar_list <- list$sugar_g
fat_list <- list$fat_g
sod_list <- list$sodium_mg

# append new values
size_list <- append(size_list, 28)
sugar_list <- append(sugar_list, 12)
fat_list <- append(fat_list, 0.5)
sod_list <- append(sod_list, 130)

# new list for normalized values
sugar_norm <- stand01(x = sugar_list / size_list)
fat_norm <- stand01(x = fat_list / size_list)
sod_norm <- stand01(x = sod_list / size_list)

# Get value
sugar_aj <- tail(sugar_norm, 1)
fat_aj <- tail(fat_norm, 1)
sod_aj <- tail(sod_norm, 1)

# Set up predict data
predict.data <- data.frame(sugar = sugar_aj, fat = fat_aj, sodium = sod_aj)


#predict (object = mod.fit.Ha, newdata = predict.data , type = "response")
model.pred <- predict(object = model_cereal_shelves_linear, newdata = predict.data, type = "pro
print("Shelf 1 Probability")
```

```
## [1] "Shelf 1 Probability"
```

```
model.pred[1]
```

```
##          1
## 0.05326849
```

```
print("Shelf 2 Probability")
```

```
## [1] "Shelf 2 Probability"
```

```
model.pred[2]
```

```
##         2
## 0.4719426
```

```
print("Shelf 3 Probability")
```

```
## [1] "Shelf 3 Probability"
```

```
model.pred[3]
```

```
##         3
## 0.2004274
```

```
print("Shelf 4 Probability")
```

```
## [1] "Shelf 4 Probability"
```

```
model.pred[4]
```

```
##         4
## 0.2743615
```

The probability of Apple Jacks being placed on each shelf is:

- Shelf 1: 5 percent
- Shelf 2: 47 percent
- Shelf 3: 20 percent
- Shelf 4: 27 percent

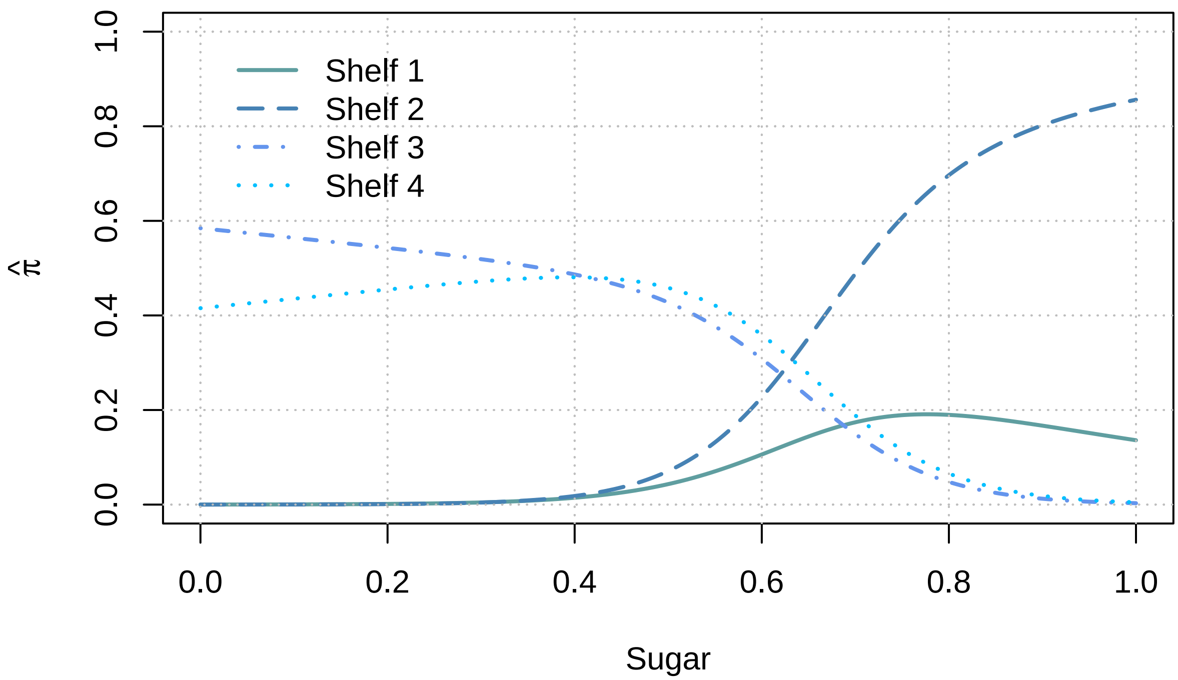
Therefore, Apple Jacks has the highest probability of being placed on shelf 2, followed by shelf 4, shelf 3 and shelf 1.

### 1.4 Figure 3.3

(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.

```
#shelf_vs_sugar_plot <- 'fill this in'

# Create prediction model to ensure probabilities are estiamted for all four shelves
predict.d <- data.frame(sodium = mean(cereal2$sodium), sugar = c(mean(cereal2$sugar), mean(cere
pi.hat <- predict(object = model_cereal_shelves_linear, newdata = predict.d, type = "probs")
#head(pi.hat)
#pi.hat[,1]

# Create plotting area first to make sure get the whole region with respect to x-axis
curve(expr = predict(object = model_cereal_shelves_linear, newdata = data.frame(sodium = mean(c
                                                                                 sugar = x,
                                                                                 fat = mean(cere
      ylab = expression(hat(pi)), xlab = "Sugar",
```

```
      xlim = c(min(cereal2$sugar), max(cereal2$sugar)),
      ylim = c(0,1), col = "#5F9EA0", lty = "solid", lwd = 2, n = 1000,
      panel.first = grid(col = "gray", lty = "dotted"))

curve(expr = predict(object = model_cereal_shelves_linear, newdata = data.frame(sodium = mean(c
                                                                                sugar = x,
                                                                                fat = mean(cere
      ylab = expression(hat(pi)), xlab = "Sugar",
      xlim = c(min(cereal2$sugar), max(cereal2$sugar)),
      ylim = c(0,1), col = "#4682B4", lty = "longdash", lwd = 2, n = 1000,
      add = TRUE, panel.first = grid(col = "gray", lty = "dotted"))


curve(expr = predict(object = model_cereal_shelves_linear, newdata = data.frame(sodium = mean(c
                                                                                sugar = x,
                                                                                fat = mean(cere
      ylab = expression(hat(pi)), xlab = "Sugar",
      xlim = c(min(cereal2$sugar), max(cereal2$sugar)),
      ylim = c(0,1), col = "#6495ED", lty = "dotdash", lwd = 2, n = 1000,
      add = TRUE, panel.first = grid(col = "gray", lty = "dotted"))

curve(expr = predict(object = model_cereal_shelves_linear, newdata = data.frame(sodium = mean(c
                                                                                sugar = x,
                                                                                fat = mean(cere
      ylab = expression(hat(pi)), xlab = "Sugar",
      xlim = c(min(cereal2$sugar), max(cereal2$sugar)),
      ylim = c(0,1),col = "#00BFFF", lty = "dotted", lwd = 2, n = 1000,
      add = TRUE, panel.first = grid(col = "gray", lty = "dotted"))


legend(x = .01, y = 1, legend=c("Shelf 1", "Shelf 2", "Shelf 3", "Shelf 4"), lty=c("solid","lon
       col=c("#5F9EA0","#4682B4","#6495ED", "#00BFFF"), bty="n", lwd = c(2,2,2,2))
```




The plot shows that the at low levels of sugar, there is higher probability that cereals will be placed on shelves 3 and 4. However, at higher levels of sugar, there is a higher probability of the cereal being placed on shelf 2. Finally, with increases to the the level of sugar, there is a slight increase in the probability that the cereal will be placed on shelf 1, although the probability starts to decline at the highest levels of sugar.

### 1.5 Odds ratios

(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.

```
# Find the standard deviation for each of the coefficients: sodium, sugar and fat:
sd.cereal <- apply(X = cereal2[, -c(1)], MARGIN = 2, FUN = sd)
c.value <- c(sd.cereal[3], sd.cereal[1], sd.cereal[2]) # reset values so they match coefficien
print("Standard Deviations of the Model Coefficients")
```

```
## [1] "Standard Deviations of the Model Coefficients"
```

```
c.value
```

```
##    sodium     sugar       fat
## 0.2298359 0.2692078 0.2990292
```

```
# compute CI for the coefficients
conf.beta <- confint(object = model_cereal_shelves_linear, level = 0.95)
exp(conf.beta)
```

```
## , , 2
##
##                    2.5 %       97.5 %
## (Intercept) 2.984337e-03 3.303929e+08
```

```
## sodium      2.298848e-14 2.777354e-02
## sugar       7.405940e-04 2.948434e+05
## fat         6.329157e-01 5.360628e+03
##
## , , 3
##
##                    2.5 %       97.5 %
## (Intercept) 1.184664e+03 5.728017e+15
## sodium      1.879520e-18 1.071292e-04
## sugar       3.418478e-10 7.163096e-02
## fat         5.039904e-03 6.511231e+01
##
## , , 4
##
##                    2.5 %       97.5 %
## (Intercept) 8.253155e+02 3.751459e+15
## sodium      2.640256e-18 1.402564e-04
## sugar       8.040525e-10 1.578574e-01
## fat         3.752910e-03 4.675812e+01
```

```
# Construct beta hats for various shelf values
beta.hat2 <- coefficients(model_cereal_shelves_linear)[1, 2:4] # Shelf 2 Vs. Shelf 1
beta.hat3 <- coefficients(model_cereal_shelves_linear)[2, 2:4] # Shelf 1 Vs. Shelf 3
beta.hat4 <- coefficients(model_cereal_shelves_linear)[3, 2:4] # Shelf 1 Vs. Shelf 4

# Odds Ratio for Shelf = 2 vs. Shelf = 1
print("Shelf = 2 vs. Shelf = 1: Odds Ratios")
```

```
## [1] "Shelf = 2 vs. Shelf = 1: Odds Ratios"
```

```
round(exp(c.value * beta.hat2 ), 2)
```

```
## sodium  sugar    fat
##   0.02   2.06   3.37
```

```
round(1/exp(c.value * beta.hat2 ), 2)
```

```
## sodium  sugar    fat
##  55.74   0.48   0.30
```

```
# Estimate the confidence intervals for Shelf 2 vs. Shelf 1
#print("Shelf = 2 vs. Shelf = 1: Confidence Interval")
ci.OR2 <- exp(c.value * conf.beta [2:4 ,1:2 ,1])
round(data.frame (low = ci.OR2[,1], up = ci.OR2[ ,2]) , 3)
```

```
##          low     up
## sodium 0.001  0.439
## sugar  0.144 29.679
## fat    0.872 13.036
```

```r
#round(data.frame(low = 1/ci.OR2[,2], up = 1/ci.OR2[ ,1]), 3)

# Odds Ratio for Shelf = 3 vs. Shelf = 1: Odds Ratios")
print("Shelf = 3 vs. Shelf = 1")
```

```
## [1] "Shelf = 3 vs. Shelf = 1"
```

```r
round(exp(c.value * beta.hat3 ), 3)
```

```
## sodium  sugar    fat
##  0.003  0.037  0.847
```

```r
round(1/exp(c.value * beta.hat3 ), 3)
```

```
##  sodium   sugar     fat
## 311.361  26.810   1.181
```

```r
#print("Shelf = 3 vs. Shelf = 1: Confidence Interval")
ci.OR3 <- exp(c.value * conf.beta [2:4 ,1:2 ,2])
round(data.frame (low = ci.OR3[,1], up = ci.OR3[ ,2]) , 4)
```

```
##           low     up
## sodium 0.0001 0.1223
## sugar  0.0028 0.4918
## fat    0.2056 3.4861
```

```r
#round(data.frame(low = 1/ci.OR3[,2], up = 1/ci.OR3[ ,1]), 4)


# Odds Ratio for Shelf = 4 vs. Shelf = 1: Odds Ratios")
print("Shelf = 4 vs. Shelf = 1")
```

```
## [1] "Shelf = 4 vs. Shelf = 1"
```

```r
round(exp(c.value * beta.hat4 ), 3)
```

```
## sodium  sugar    fat
##  0.003  0.047  0.771
```

```r
round(1/exp(c.value * beta.hat4 ), 3)
```

```
##  sodium   sugar     fat
## 290.306  21.483   1.297
```

```r
#print("Shelf = 4 vs. Shelf = 1: Confidence Interval")
ci.OR4 <- exp(c.value * conf.beta [2:4 ,1:2 ,3])
round(data.frame (low = ci.OR4[,1], up = ci.OR4[ ,2]) , 4)
```

```
##           low     up
## sodium 0.0001 0.1301
## sugar  0.0036 0.6084
## fat    0.1882 3.1574
```

```r
#round(data.frame(low = 1/ci.OR4[,2], up = 1/ci.OR4[ ,1]), 4)
```

Summary findings for the odd ratios can be found below:

- The estimated odds of a cereal being on shelf 2 vs. shelf 1 change by 0.02 times for a 0.23 increase in sodium holding the other variables constant. Equivalently, we can say that the estimated odds of a cereal being placed on shelf 2 vs. shelf 1 change by 55.74 times for a 0.23 decrease in sodium holding the other variables constant. The estimated odds of a cereal being on shelf 2 vs. shelf 1 change by 2.06 times for a 0.27 increase in sugar holding the other variables constant. Equivalently, we can say that the estimated odds of a cereal being placed on shelf 2 vs. shelf 1 change by 0.48 times for a 0.27 decrease in sugar holding the other variables constant. The estimated odds of a cereal being on shelf 2 vs. shelf 1 change by 3.37 times for a 0.30 (rounded) increase in fat holding the other variables constant. Equivalently, we can say that the estimated odds of a cereal being placed on shelf 2 vs. shelf 1 change by 0.30 times for a 0.30 (rounded) decrease in sugar holding the other variables constant.

*With 95% confidence, the odds of shelf 2 instead of shelf 1 change by 0.001 to 0.439 times when sodium changes by 0.23, while holding the other variables constant. With 95% confidence, the odds of level 2 instead of level 1 change by 0.14 to 29.68 times when sugar changes by 0.27, while holding the other variables constant. With 95% confidence, the odds of level 2 instead of level 1 change by 0.87 to 13.04 times when fat changes by 0.30, while holding the other variables constant.

- The estimated odds of a cereal being on shelf 3 vs. shelf 1 change by 0.003 times for a 0.23 increase in sodium holding the other variables constant. Equivalently, we can say that the estimated odds of a cereal being placed on shelf 3 vs. shelf 1 change by 311.36 times for a 0.23 decrease in sodium holding the other variables constant. The estimated odds of a cereal being on shelf 3 vs. shelf 1 change by 0.04 times for a 0.27 increase in sugar holding the other variables constant. Equivalently, we can say that the estimated odds of a cereal being placed on shelf 3 vs. shelf 1 change by 26.81 times for a 0.27 decrease in sugar holding the other variables constant. The estimated odds of a cereal being on shelf 3 vs. shelf 1 change by 0.85 times for a 0.30 (rounded) increase in fat holding the other variables constant. Equivalently, we can say that the estimated odds of a cereal being placed on shelf 3 vs. shelf 1 change by 1.18 times for a 0.30 (rounded) decrease in sugar holding the other variables constant.

*With 95% confidence, the odds of shelf 3 instead of shelf 1 change by 0.0001 to 0.1223 times when sodium changes by 0.23, while holding the other variables constant. With 95% confidence, the odds of level 3 instead of level 1 change by 0.0028 to 0.49 times when sugar changes by 0.27, while holding the other variables constant. With 95% confidence, the odds of level 3 instead of level 1 change by 0.21 to 3.49 times when fat changes by 0.30, while holding the other variables constant.

- The estimated odds of a cereal being on shelf 4 vs. shelf 1 change by 0.003 times for a 0.23 increase in sodium holding the other variables constant. Equivalently, we can say that the estimated odds of a cereal being placed on shelf 4 vs. shelf 1 change by 290.31 times for a 0.23 decrease in sodium holding the other variables constant. The estimated odds of a cereal being on shelf 4 vs. shelf 1 change by 0.05 times for a 0.27 increase in sugar holding the other variables constant. Equivalently, we can say that the estimated odds of a cereal being placed on shelf 4 vs. shelf 1 change by 21.48 times for a 0.27 decrease in sugar holding the other variables constant. The estimated odds of a cereal being on shelf 4 vs. shelf 1 change by 0.77 times for a 0.30 (rounded) increase in fat holding the other variables constant. Equivalently, we can say that the estimated odds of a cereal being placed on shelf 4 vs. shelf 1 change by

1.30 times for a 0.30 (rounded) decrease in sugar holding the other variables constant.

*With 95% confidence, the odds of shelf 4 instead of shelf 1 change by 0.0001 to 0.1301 times when sodium changes by 0.23, while holding the other variables constant. With 95% confidence, the odds of level 4 instead of level 1 change by 0.0036 to 0.6084 times when sugar changes by 0.27, while holding the other variables constant. With 95% confidence, the odds of level 3 instead of level 1 change by 0.19 to 3.16 times when fat changes by 0.30, while holding the other variables constant.

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

Read the example **'Alcohol Consumption'** in chapter 4.2.2 of the textbook(Bilder and Loughin's "Analysis of Categorical Data with R). 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.*

```r
drinks <- read.csv('DeHartSimplified.csv')
```

## 2.1 EDA

(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.

> 'Fill this in: What do you learn?'

## 2.2 Hypothesis One

(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.

> 'Fill this in: What do you learn?'

## 2.3 Hypothesis Two

(1 point) 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.

'Fill this in: What do you learn?'