

W271 Assignment 2

Ren Tu

We start by loading the libraries required for completing this assignment:

```
### Load libraries for assignment
library(ggplot2)
library(GGally)

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2

library(reshape2)
library(nnet)
library(car)

## Loading required package: carData
```

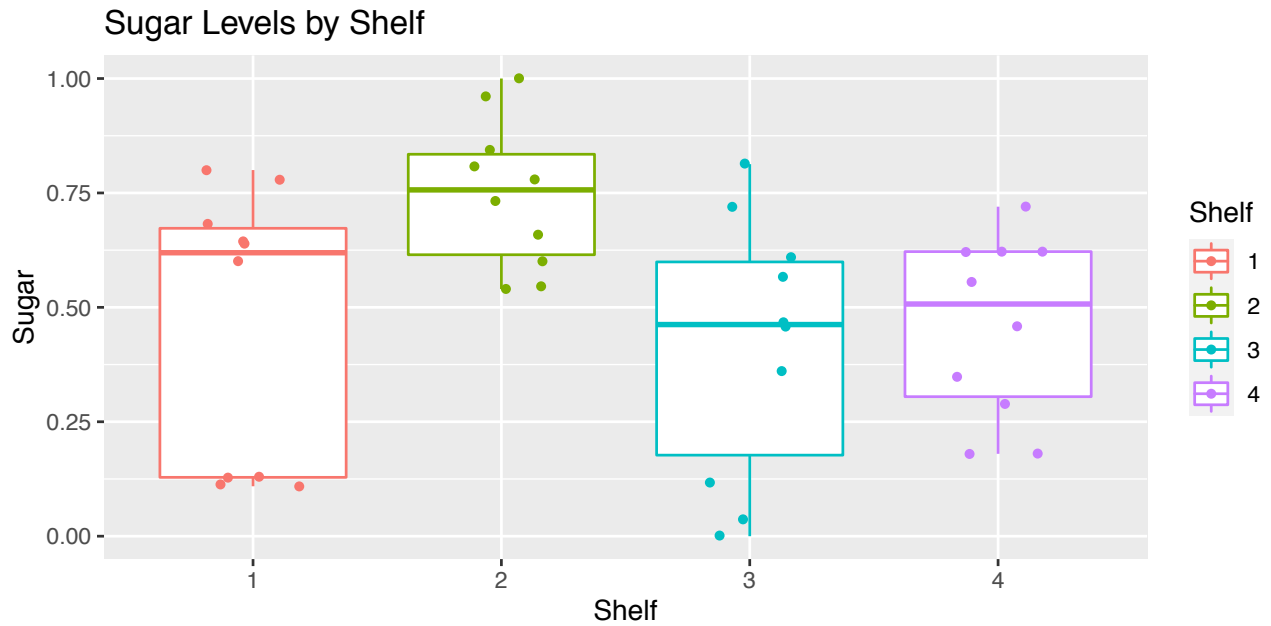
Question 1

Question 1.1: Explanatory Variable Plots

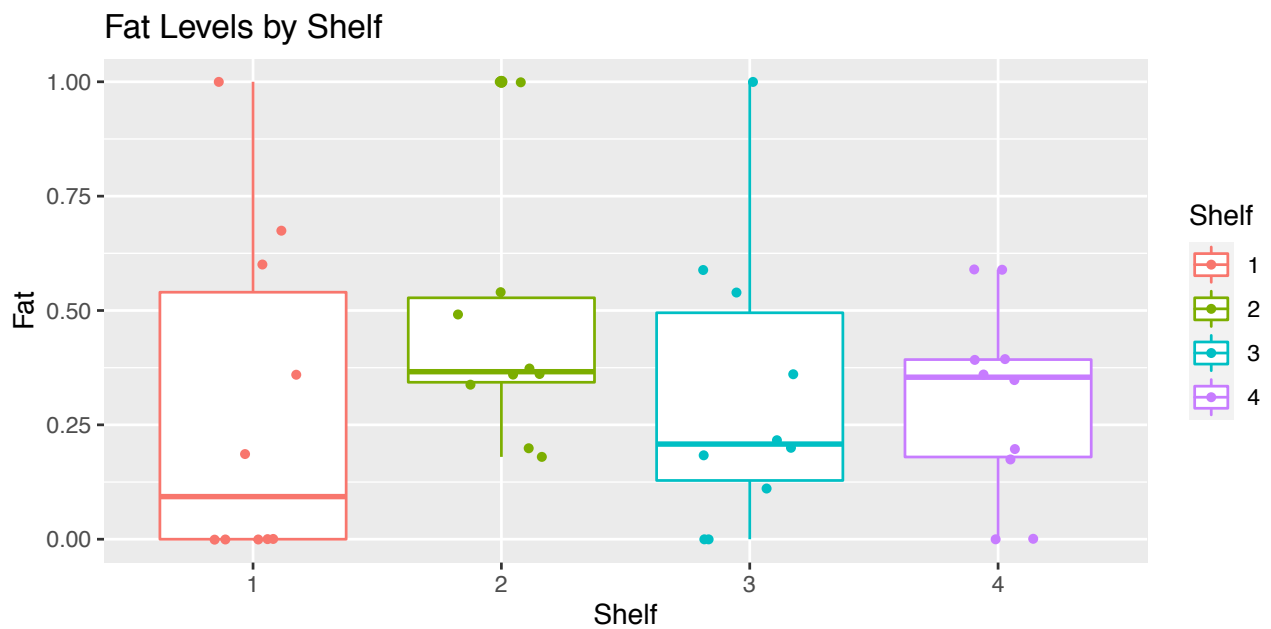
We will begin by reformatting the variables in the cereal dataset and exploring the relationships to shelf number.

```
### Read in and reformat the cereal dataset
data1 <- read.csv("cereal_dillons.csv")
reformat <- function(x) { (x - min(x)) / (max(x) - min(x)) }
cereal <- data.frame(Shelf = as.factor(data1$Shelf),
                    sugar = reformat(data1$sugar_g/data1$size_g),
                    fat = reformat(data1$fat_g/data1$size_g),
                    sodium = reformat(data1$sodium_mg/data1$size_g))

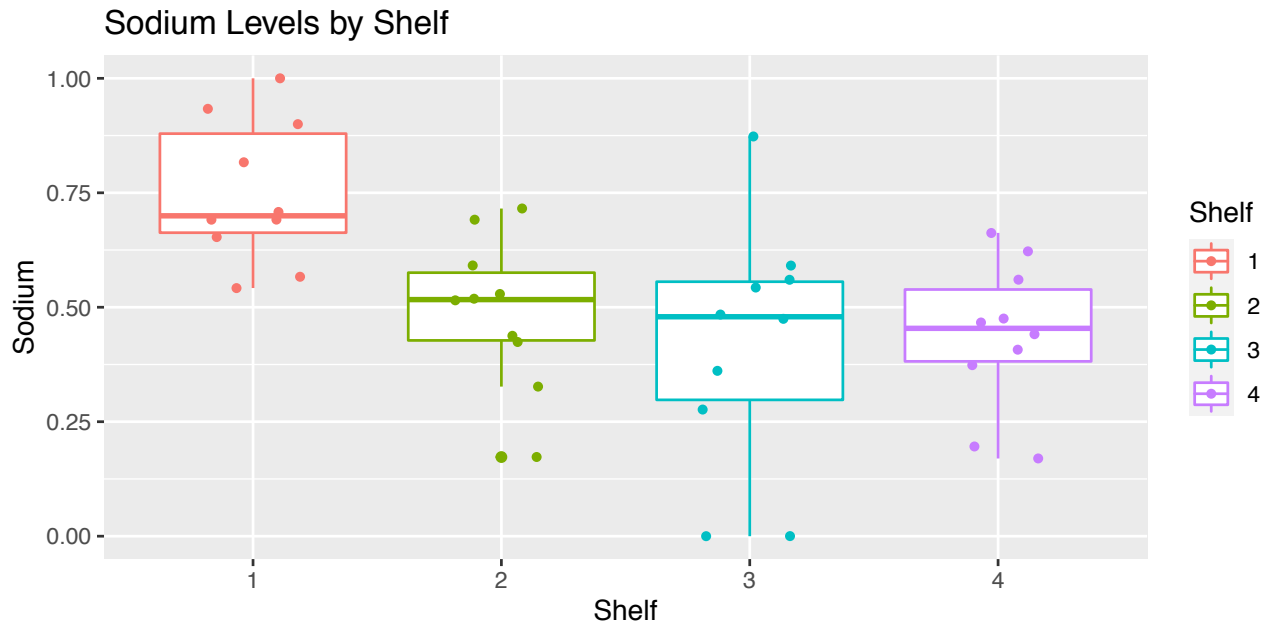
### Boxplot with dot plot overlay for sugar and shelf
plot1 <- ggplot(cereal, aes(x=Shelf, y=sugar, color=Shelf))
plot1 <- plot1 + geom_boxplot() + geom_jitter(shape=16, position=position_jitter(0.2))
plot1 <- plot1 + labs(title="Sugar Levels by Shelf", x="Shelf", y="Sugar")
plot1
```



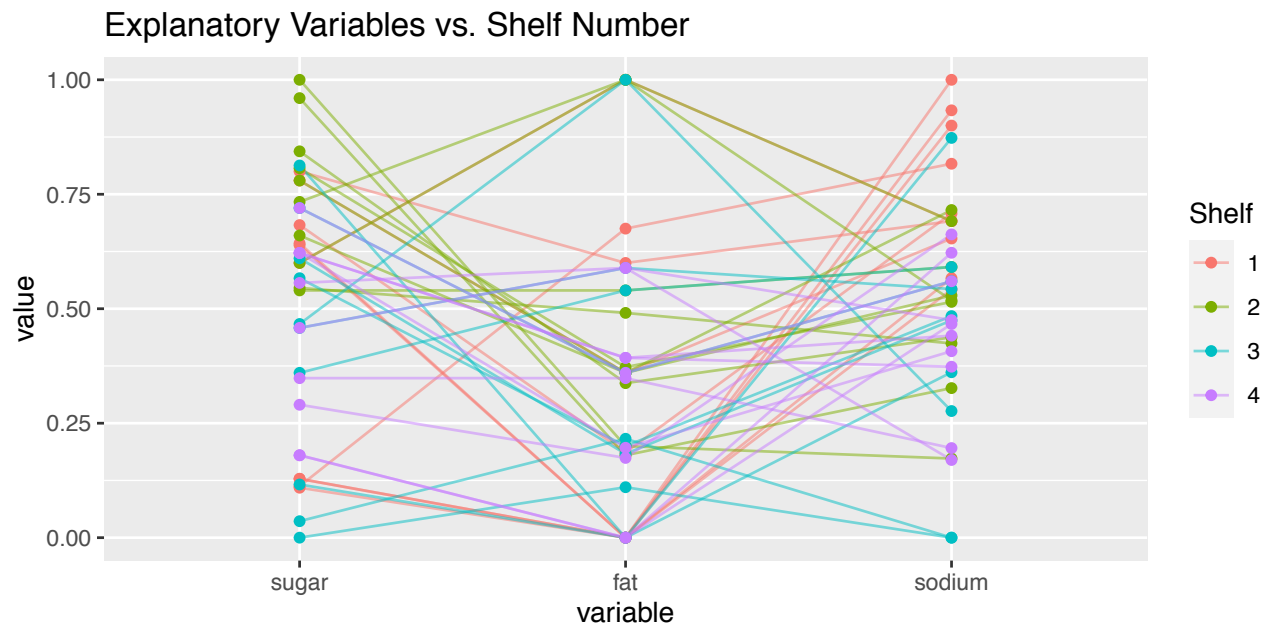
```
### Boxplot with dot plot overlay for fat and shelf
plot2 <- ggplot(cereal, aes(x=Shelf, y=fat, color=Shelf))
plot2 <- plot2 + geom_boxplot() + geom_jitter(shape=16, position=position_jitter(0.2))
plot2 <- plot2 + labs(title="Fat Levels by Shelf", x="Shelf", y="Fat")
plot2
```



```
### Boxplot with dot plot overlay for sodium and shelf
plot3 <- ggplot(cereal, aes(x=Shelf, y=sodium, color=Shelf))
plot3 <- plot3 + geom_boxplot() + geom_jitter(shape=16, position=position_jitter(0.2))
plot3 <- plot3 + labs(title="Sodium Levels by Shelf", x="Shelf", y="Sodium")
plot3
```



```
### Parallel coordinates plot for explanatory variables and shelf number
plot4 <- ggparcoord(data=cereal, columns=2:4, groupColumn=1, showPoints=TRUE,
  alphaLines=0.5, scale="globalminmax",
  title="Explanatory Variables vs. Shelf Number")
plot4
```



With the reformatting of the variables, the sugar, fat and sodium content are now on a 0 to 1 scale. From the sugar boxplot, shelf 2 cereals tend to have above average sugar content with sugar values above 0.5 while other shelves have wide dispersion in sugar content. From the fat boxplot, shelf 4 is the only shelf without cereals with the highest fat levels, but all shelves tend to have a wide range of fat content. From the sodium boxplot, shelf 1 stands out as having only high-sodium cereals. The parallel coordinates plot confirm the boxplot findings with high concentrations of high-sugar cereals for shelf 2 and high-sodium cereals for shelf 1. Shelf 4 is also notable for not having high content levels for sugar, fat and sodium. Shelf 3 has no discernible

pattern with cereals that cover approximately the entire range of sugar, fat and sodium levels.

Question 1.2: Multinomial Regression Model

With the shelf numbers listed from bottom to top with the integers from 1 to 4, we can consider whether to take ordinality into account for modeling. Ordinality makes sense when a variable has a natural scale that is relevant to the research question. While the shelf variable in this dataset does have a natural scale from a height perspective, it is not directly related the research question that is focused on the attractiveness of shelves to draw customer attention. Without further details on the exact height levels of each shelf, we cannot assume that the higher shelves always draw more or less attention or vice versa. Therefore, we will estimate a multinomial model that does not take ordinality of shelves into account.

```
### Estimate multinomial regression and check convergence
q1.model1 <- multinom(formula = Shelf ~ sugar+fat+sodium, data=cereal,
                      trace=FALSE)
print(paste("Convergence value is: ", q1.model1$convergence))
```

```
## [1] "Convergence value is: 0"
```

```
summary(q1.model1)
```

```
## Call:
## multinom(formula = Shelf ~ sugar + fat + sodium, data = cereal,
##          trace = FALSE)
##
## 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
```

The model did indeed converge with a convergence value of zero. Our estimated multinomial model can be expressed as the following:

$$\begin{aligned} \log(\hat{\pi}_2/\hat{\pi}_1) &= 6.90 + 2.69sugar + 4.06fat - 17.49sodium \\ \log(\hat{\pi}_3/\hat{\pi}_1) &= 21.68 - 12.22sugar - 0.56fat - 24.98sodium \\ \log(\hat{\pi}_4/\hat{\pi}_1) &= 21.29 - 11.39sugar - 0.87fat - 24.67sodium \end{aligned}$$

Next, we will perform LRTs to evaluate variable importance at a 5% level of significance:

```
### LRT of multinomial regression model
Anova(q1.model1)
```

```
## 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
```

The LRT results indicate that sugar and sodium are statistically significant variables with p-values below 0.001. Therefore sugar and sodium levels have sufficient evidence for importance in explaining shelf number given the presence of other variables in the model. On the other hand, fat levels do not have sufficient evidence for importance with a p-value of 0.15.

Next, we will examine whether there are any significant variable interactions at a 5% level of significance.

```
### Estimate multinomial model with interactions and check convergence
q1.model2 <- multinom(formula = Shelf ~ (sugar+fat+sodium)^3, data=cereal,
                      maxit=3000, trace=FALSE)
print(paste("Convergence value is: ", q1.model2$convergence))
```

```
## [1] "Convergence value is:  0"
```

```
### LRT of model with interactions
Anova(q1.model2)
```

```
## 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 4.7772  3  0.1888585
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The model with interactions did indeed converge with a convergence value of zero. With p-values of 0.36, 0.39, 0.37 and 0.19, none of the four interaction terms are significant. Therefore we will use the first multinomial model with no interactions as our final model.

Question 1.3: Apple Jacks

Here we will estimate the shelf probabilities for Kellogg's Apple Jacks given its characteristics: 28g serving size, 12g sugar, 0.5g fat, and 130mg sodium

```
### Function to rescale values using same methodology as original dataset
newdata.scale <- function(n, orig) { (n - min(orig)) / (max(orig) - min(orig)) }

### Convert sugar, fat and sodium for Apple Jacks to new scale
newdata.sugar <- newdata.scale(12/28, data1$sugar_g/data1$size_g)
newdata.fat <- newdata.scale(0.5/28, data1$fat_g/data1$size_g)
newdata.sodium <- newdata.scale(130/28, data1$sodium_mg/data1$size_g)

### Predict shelf probabilities for Apple Jacks
applejacks <- data.frame(rbind(c(0,newdata.sugar,newdata.fat,newdata.sodium)))
colnames(applejacks) <- c("Shelf","sugar","fat","sodium")
```

```
pi.hat.q1.3 <- predict(object=q1.model1, newdata=applejacks, type="probs")
pi.hat.q1.3
```

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

Our multinomial model predicts that Apple Jacks has a 5.32% chance of being on shelf 1, 47.19% chance of being on shelf 2, 20.04% chance of being on shelf 3, and 27.44% chance of being on shelf 4.

Question 1.4: Sugar Content Plot

Using our final multinomial model, we will construct a plot with estimated probability for a shelf on the y-axis and sugar content on the x-axis. We will use mean fat and sodium content from the original dataset as inputs in the model.

```
### Create sequence of values from 0 to 1 for sugar content
sugar.vals <- seq(0, 1, 0.01)

### Create sequence of mean fat and sodium content values
fat.vals <- rep(mean(cereal$fat), 101)
sodium.vals <- rep(mean(cereal$sodium), 101)

### Create sequence of dummy values for Shelf
shelf.vals <- rep(0, 101)

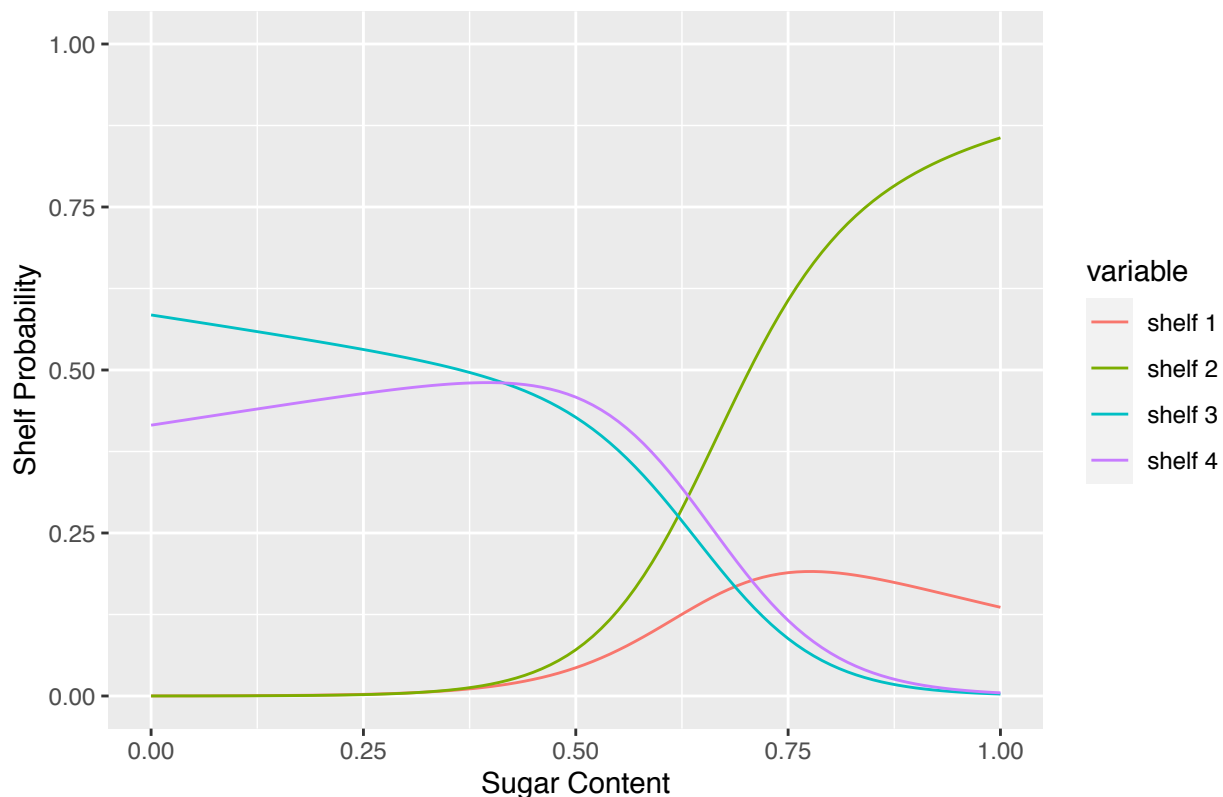
### Combine sequences into dataframe and feed into model
sugar.df <- data.frame(shelf.vals, sugar.vals, fat.vals, sodium.vals)
colnames(sugar.df) <- c("Shelf", "sugar", "fat", "sodium")
pi.hat.q1.4 <- predict(object=q1.model1, newdata=sugar.df, type="probs")

### Transform model output into plot-friendly format
probplot <- data.frame(cbind(pi.hat.q1.4, sugar.vals))
colnames(probplot) <- c("shelf 1", "shelf 2", "shelf 3", "shelf 4", "sugar")
probplot2 <- melt(probplot, id="sugar")

### Plot probabilities for the four shelves
sugarplot <- ggplot(probplot2, aes(x=sugar, y=value, color=variable))
sugarplot <- sugarplot + geom_line() + ylim(0,1)
sugarplot <- sugarplot + labs(title="Shelf Probabilities From Sugar Content",
                             x="Sugar Content", y="Shelf Probability")

sugarplot
```

Shelf Probabilities From Sugar Content



```
### Get original sugar content values to interpret results
minratio <- min(data1$sugar_g/data1$size_g)
maxratio <- max(data1$sugar_g/data1$size_g)
print(paste("Lowest sugar to serving size ratio is: ", minratio))
```

```
## [1] "Lowest sugar to serving size ratio is: 0"
```

```
print(paste("Highest sugar to serving size ratio is: ", maxratio))
```

```
## [1] "Highest sugar to serving size ratio is: 0.555555555555556"
```

We can interpret the shelf probability plot as follows with respect to sugar content when holding fat and sodium constant at their mean levels: In the lower half of the sugar content range, shelves 3 and 4 each have about 50% probability while shelf 1 and 2 stay near 0%. Then as sugar content increases in the top half of its range, the probability for shelf 2 increases rapidly towards above 80%, shelf 1 increases gradually to 15%, while shelf 3 and 4 decrease toward 0%. In other words, low-sugar cereals tend to be placed on shelves 3 and 4 while high-sugar cereals tend to be placed on shelf 2.

One important limitation to the shelf probability plot is the sugar content range used to estimate the model. As we can see from the original dataset values, the sugar to serving size ratio ranged from 0 to 0.5556 before being reformatted to the 0 to 1 scale. Therefore, the above plot is only applicable to cereals with sugar to serving size ratios less than 0.5556 and our interpretations refer to this specific range only.

Question 1.5: Odds Ratios and Confidence Intervals

We can use the following rules when calculating odds ratios for each variable from our multinomial model (Bilder textbook page 160):

Odds of category j response vs. category 1 response change by $\exp(c\beta_{jr})$ times for every c-unit increase in variable x_r
Odds of category j response vs. category j' response change by $\exp(c(\beta_{jr} - \beta_{j'r}))$ times for every c-unit increase in variable x_r

Since our explanatory variables are based on a 0 to 1 scale, using 0.25 as a c-unit increase covers a large enough portion of the scale to give us both practically significant and interpretable odds ratios.

Now we will calculate the odds ratios for a 0.25-increase for each variable for every combination of shelves:

```
### Get model beta coefficients
shelf.2vs1.betas <- coefficients(q1.model1)[1, 2:4]
shelf.3vs1.betas <- coefficients(q1.model1)[2, 2:4]
shelf.4vs1.betas <- coefficients(q1.model1)[3, 2:4]

### Calculate odds ratios for shelves 2/3/4 vs. shelf 1
c <- 0.25
shelf.2vs1.odds <- round(exp(c * shelf.2vs1.betas), 4)
shelf.3vs1.odds <- round(exp(c * shelf.3vs1.betas), 4)
shelf.4vs1.odds <- round(exp(c * shelf.4vs1.betas), 4)

### Calculate odds ratios for 3vs2, 4vs2 and 4vs3
shelf.3vs2.odds <- round(exp(c * (shelf.3vs1.betas - shelf.2vs1.betas)), 4)
shelf.4vs2.odds <- round(exp(c * (shelf.4vs1.betas - shelf.2vs1.betas)), 4)
shelf.4vs3.odds <- round(exp(c * (shelf.4vs1.betas - shelf.3vs1.betas)), 4)

q1.odds.df <- data.frame(rbind(shelf.2vs1.odds, shelf.3vs1.odds,
                              shelf.4vs1.odds, shelf.3vs2.odds,
                              shelf.4vs2.odds, shelf.4vs3.odds))

q1.odds.df

##           sugar    fat sodium
## shelf.2vs1.odds 1.9606 2.7626 0.0126
## shelf.3vs1.odds 0.0472 0.8700 0.0019
## shelf.4vs1.odds 0.0579 0.8045 0.0021
## shelf.3vs2.odds 0.0241 0.3149 0.1539
## shelf.4vs2.odds 0.0295 0.2912 0.1661
## shelf.4vs3.odds 1.2284 0.9247 1.0791
```

Next, we will calculate the 95% Wald confidence intervals around each odds ratio. We can use the following equations (Bilder textbook page 83):

Wald C.I. for odds of category j response vs. category 1 response for c-unit change:

$$\exp(c\beta_{jr} \pm cZ_{1-\alpha/2}\sqrt{\text{Var}(\beta_{jr})})$$

Wald C.I. for odds of category j response vs. category j' response for c-unit change:

$$\begin{aligned} &\exp[c(\beta_{jr} - \beta_{j'r}) \pm cZ_{1-\alpha/2}\sqrt{\text{Var}(\beta_{jr} - \beta_{j'r})}] \\ &= \exp[c(\beta_{jr} - \beta_{j'r}) \pm cZ_{1-\alpha/2}\sqrt{\text{Var}(\beta_{jr}) + \text{Var}(\beta_{j'r}) - 2\text{Cov}(\beta_{jr}, \beta_{j'r})}] \end{aligned}$$


```

### Wald 95% intervals for 2vs1, 3vs1 and 4vs1
c <- 0.25
betas.base1.sug <- coefficients(q1.model1)[1:3, 2]
betas.base1.fat <- coefficients(q1.model1)[1:3, 3]
betas.base1.sod <- coefficients(q1.model1)[1:3, 4]
vcov1 <- vcov(q1.model1)
var.base1.sug <- c(vcov1[2,2], vcov1[6,6], vcov1[10,10])
var.base1.fat <- c(vcov1[3,3], vcov1[7,7], vcov1[11,11])
var.base1.sod <- c(vcov1[4,4], vcov1[8,8], vcov1[12,12])
CI.base1.sug.lo <- exp(c*betas.base1.sug + c*qnrm(p=0.025)*sqrt(var.base1.sug))
CI.base1.sug.hi <- exp(c*betas.base1.sug + c*qnrm(p=0.975)*sqrt(var.base1.sug))
CI.base1.fat.lo <- exp(c*betas.base1.fat + c*qnrm(p=0.025)*sqrt(var.base1.fat))
CI.base1.fat.hi <- exp(c*betas.base1.fat + c*qnrm(p=0.975)*sqrt(var.base1.fat))
CI.base1.sod.lo <- exp(c*betas.base1.sod + c*qnrm(p=0.025)*sqrt(var.base1.sod))
CI.base1.sod.hi <- exp(c*betas.base1.sod + c*qnrm(p=0.975)*sqrt(var.base1.sod))

### Wald 95% intervals for sugar for 3vs2, 4vs2, 4vs3
cov.sug <- c(vcov1[6,2], vcov1[10,2], vcov1[10,6])
jminusj.sug <- c(betas.base1.sug[2]-betas.base1.sug[1],
                betas.base1.sug[3]-betas.base1.sug[1],
                betas.base1.sug[3]-betas.base1.sug[2])
var.sug.jplusj <- c(var.base1.sug[2]+var.base1.sug[1],
                    var.base1.sug[3]+var.base1.sug[1],
                    var.base1.sug[3]+var.base1.sug[2])
sqrt.term.sug <- sqrt(var.sug.jplusj-2*cov.sug)
CI.base23.sug.lo <- exp(c*jminusj.sug + c*qnrm(p=0.025)*sqrt.term.sug)
CI.base23.sug.hi <- exp(c*jminusj.sug + c*qnrm(p=0.975)*sqrt.term.sug)

### Wald 95% intervals for fat for 3vs2, 4vs2, 4vs3
cov.fat <- c(vcov1[7,3], vcov1[11,3], vcov1[11,7])
jminusj.fat <- c(betas.base1.fat[2]-betas.base1.fat[1],
                betas.base1.fat[3]-betas.base1.fat[1],
                betas.base1.fat[3]-betas.base1.fat[2])
var.fat.jplusj <- c(var.base1.fat[2]+var.base1.fat[1],
                    var.base1.fat[3]+var.base1.fat[1],
                    var.base1.fat[3]+var.base1.fat[2])
sqrt.term.fat <- sqrt(var.fat.jplusj-2*cov.fat)
CI.base23.fat.lo <- exp(c*jminusj.fat + c*qnrm(p=0.025)*sqrt.term.fat)
CI.base23.fat.hi <- exp(c*jminusj.fat + c*qnrm(p=0.975)*sqrt.term.fat)

### Wald 95% intervals for sodium for 3vs2, 4vs2, 4vs3
cov.sod <- c(vcov1[8,4], vcov1[12,4], vcov1[12,8])
jminusj.sod <- c(betas.base1.sod[2]-betas.base1.sod[1],
                betas.base1.sod[3]-betas.base1.sod[1],
                betas.base1.sod[3]-betas.base1.sod[2])
var.sod.jplusj <- c(var.base1.sod[2]+var.base1.sod[1],
                    var.base1.sod[3]+var.base1.sod[1],
                    var.base1.sod[3]+var.base1.sod[2])
sqrt.term.sod <- sqrt(var.sod.jplusj-2*cov.sod)
CI.base23.sod.lo <- exp(c*jminusj.sod + c*qnrm(p=0.025)*sqrt.term.sod)
CI.base23.sod.hi <- exp(c*jminusj.sod + c*qnrm(p=0.975)*sqrt.term.sod)

### Round all results to 4 decimals

```

```

CI.base1.sug.lo <- round(CI.base1.sug.lo, 4)
CI.base1.sug.hi <- round(CI.base1.sug.hi, 4)
CI.base1.fat.lo <- round(CI.base1.fat.lo, 4)
CI.base1.fat.hi <- round(CI.base1.fat.hi, 4)
CI.base1.sod.lo <- round(CI.base1.sod.lo, 4)
CI.base1.sod.hi <- round(CI.base1.sod.hi, 4)
CI.base23.sug.lo <- round(CI.base23.sug.lo, 4)
CI.base23.sug.hi <- round(CI.base23.sug.hi, 4)
CI.base23.fat.lo <- round(CI.base23.fat.lo, 4)
CI.base23.fat.hi <- round(CI.base23.fat.hi, 4)
CI.base23.sod.lo <- round(CI.base23.sod.lo, 4)
CI.base23.sod.hi <- round(CI.base23.sod.hi, 4)

### Combine sugar results into dataframe table
df.sug <- data.frame(rbind(cbind(CI.base1.sug.lo, CI.base1.sug.hi),
                              cbind(CI.base23.sug.lo, CI.base23.sug.hi)))
rownames(df.sug) <- c("Sugar Shelf 2 vs. 1", "Sugar Shelf 3 vs. 1",
                     "Sugar Shelf 4 vs. 1", "Sugar Shelf 3 vs. 2",
                     "Sugar Shelf 4 vs. 2", "Sugar Shelf 4 vs. 3")
df.sug$orig <- q1.odds.df$sugar
colnames(df.sug) <- c("Lower Bound", "Upper Bound", "Original Estimate")

### Combine fat results into dataframe table
df.fat <- data.frame(rbind(cbind(CI.base1.fat.lo, CI.base1.fat.hi),
                              cbind(CI.base23.fat.lo, CI.base23.fat.hi)))
rownames(df.fat) <- c("Fat Shelf 2 vs. 1", "Fat Shelf 3 vs. 1",
                     "Fat Shelf 4 vs. 1", "Fat Shelf 3 vs. 2",
                     "Fat Shelf 4 vs. 2", "Fat Shelf 4 vs. 3")
df.fat$orig <- q1.odds.df$fat
colnames(df.fat) <- c("Lower Bound", "Upper Bound", "Original Estimate")

### Combine sodium results into dataframe table
df.sod <- data.frame(rbind(cbind(CI.base1.sod.lo, CI.base1.sod.hi),
                              cbind(CI.base23.sod.lo, CI.base23.sod.hi)))
rownames(df.sod) <- c("Sodium Shelf 2 vs. 1", "Sodium Shelf 3 vs. 1",
                     "Sodium Shelf 4 vs. 1", "Sodium Shelf 3 vs. 2",
                     "Sodium Shelf 4 vs. 2", "Sodium Shelf 4 vs. 3")
df.sod$orig <- q1.odds.df$sodium
colnames(df.sod) <- c("Lower Bound", "Upper Bound", "Original Estimate")

```

Here are the 95% Wald confidence interval results alongside the original odds ratio estimates for a 0.25 increase in each of the explanatory variables:

Display results

df.sug

##		Lower Bound	Upper Bound	Original Estimate
##	Sugar Shelf 2 vs. 1	0.1650	23.3022	1.9606
##	Sugar Shelf 3 vs. 1	0.0043	0.5173	0.0472
##	Sugar Shelf 4 vs. 1	0.0053	0.6303	0.0579
##	Sugar Shelf 3 vs. 2	0.0020	0.2869	0.0241
##	Sugar Shelf 4 vs. 2	0.0026	0.3405	0.0295
##	Sugar Shelf 4 vs. 3	0.4715	3.2001	1.2284

df.fat

##		Lower Bound	Upper Bound	Original Estimate
##	Fat Shelf 2 vs. 1	0.8919	8.5567	2.7626
##	Fat Shelf 3 vs. 1	0.2664	2.8406	0.8700
##	Fat Shelf 4 vs. 1	0.2475	2.6150	0.8045
##	Fat Shelf 3 vs. 2	0.0818	1.2126	0.3149
##	Fat Shelf 4 vs. 2	0.0759	1.1176	0.2912
##	Fat Shelf 4 vs. 3	0.3917	2.1832	0.9247

df.sod

##		Lower Bound	Upper Bound	Original Estimate
##	Sodium Shelf 2 vs. 1	0.0004	0.4082	0.0126
##	Sodium Shelf 3 vs. 1	0.0000	0.1017	0.0019
##	Sodium Shelf 4 vs. 1	0.0000	0.1088	0.0021
##	Sodium Shelf 3 vs. 2	0.0101	2.3428	0.1539
##	Sodium Shelf 4 vs. 2	0.0111	2.4837	0.1661
##	Sodium Shelf 4 vs. 3	0.3754	3.1022	1.0791

To interpret the results, we can look at the Shelf 3 vs. 2 in the Sugar table as an example. The odds of cereal being placed on shelf 3 versus shelf 2 change by a factor of 0.0241 for every 0.25 increase in sugar content on the 0 to 1 scale while holding other variables constant. And with 95% confidence, the odds of cereal being placed on shelf 3 versus shelf 2 change by a factor 0.0020 to 0.2869 for every 0.25 increase in sugar content while holding other variables constant. In other words, as sugar content increases, the chance of cereal being on shelf 3 decreases significantly relative to shelf 2.

Among the sugar variable odds ratios, there are four shelf combinations with confidence intervals that do not include 1: 3 vs. 1, 4 vs. 1, 3 vs. 2 and 4 vs. 2. Each of these four combinations have intervals below 1, indicating strong evidence that the odds of being on shelves 1 and 2 increase relative to shelves 3 and 4 as sugar content increases while holding other variables constant. This is consistent with the sugar content shelf probability plot in generated in Question 1.4 in which the probabilities for shelves 3 and 4 decrease and the probability for shelves 1 and 2 increase as sugar content increases towards the right side of the plot.

Among the fat variable odds ratios, none of the shelf combinations have confidence intervals that exclude 1. Therefore, there is insufficient evidence that the odds of being on any shelf changes as fat content changes while holding other variables constant. This aligns with our earlier findings as our multinomial model found that the fat variable was not statistically significant. In addition, the boxplots generated in Question 1.1 showed that all shelves tend to have a wide range of fat content, reaffirming the odds ratios here.

Among the sodium variable odds ratios, three shelf combinations have confidence intervals that exclude 1: 2 vs. 1, 3 vs. 1 and 4 vs. 1. Each of these three combinations have intervals below 1, indicating strong evidence that the odds of being on shelf 1 increase relative to shelves 2, 3 and 4 as sodium content increases while holding other variables constant. These odds ratios are consistent with the boxplots and parallel coordinates plot from Question 1.1 in which shelf 1 stood out as having high concentrations of high-sodium cereals.

Question 2

Question 2.1: Exploratory Data Analysis

To begin, we will conduct a thorough exploratory data analysis of the alcohol consumption dataset. We will place particular emphasis on the relationships between negative romantic interactions and alcohol consumption in accordance with original researcher hypotheses.

```
### Read in alcohol consumption dataset
q2 <- read.csv("DeHartSimplified.csv")

### Look at structure of dataset
str(q2)

## 'data.frame': 623 obs. of 13 variables:
## $ id : int 1 1 1 1 1 1 1 2 2 2 ...
## $ studyday: int 1 2 3 4 5 6 7 1 2 3 ...
## $ dayweek : int 6 7 1 2 3 4 5 3 4 5 ...
## $ numall : int 9 1 1 2 2 1 4 3 4 0 ...
## $ nrel : num 1 0 1 0 1.33 ...
## $ prel : num 0 0 0 1 0.333 ...
## $ negevent: num 0.4 0.25 0.267 0.533 0.663 ...
## $ posevent: num 0.525 0.7 1 0.608 0.693 ...
## $ gender : int 2 2 2 2 2 2 2 2 2 2 ...
## $ rosn : num 3.3 3.3 3.3 3.3 3.3 3.3 3.3 3.9 3.9 3.9 ...
## $ age : num 39.5 39.5 39.5 39.5 39.5 ...
## $ desired : num 5.67 2 3 3.67 3 ...
## $ state : num 4 2.78 4.22 4.11 4.22 ...

### Statistical summary of numall and desired variables
summary(q2$numall)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
## 0.000 1.000 2.000 2.524 3.750 21.000 1

summary(q2$desired)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
## 1.000 3.333 4.667 4.465 5.667 8.000 3

### Check if numall and desired variables have integer values
unique(q2$numall)

## [1] 9 1 2 4 3 0 7 6 5 12 8 NA 18 10 11 21 14 15 13

unique(q2$desired)

## [1] 5.666667 2.000000 3.000000 3.666667 4.000000 5.000000 4.666667 NA
## [9] 4.333333 2.666667 1.666667 7.666667 5.333333 1.000000 6.000000 2.333333
## [17] 7.000000 6.666667 7.333333 3.333333 6.333333 1.333333 8.000000

### Transform desired variable by multiplying by 3
q2$desired <- q2$desired * 3

### Confirm desired now has integer values
unique(q2$desired)

## [1] 17 6 9 11 12 15 14 NA 13 8 5 23 16 3 18 7 21 20 22 10 19 4 24
```

Based on definitions from the Bilder textbook (page 260), our main dependent variables will be *numall*

(number of drinks consumed) and *desired*. While no definition for the *desired* variable was given, we will assume this is the number of drinks desired.

Potential explanatory variables are *nrel* (negative romantic events), *prel* (positive romantic events), *negevent* (negative life events), *posevent* (positive life events), *rosn* (trait self-esteem), *state* (state self-esteem), and *age* (person's age).

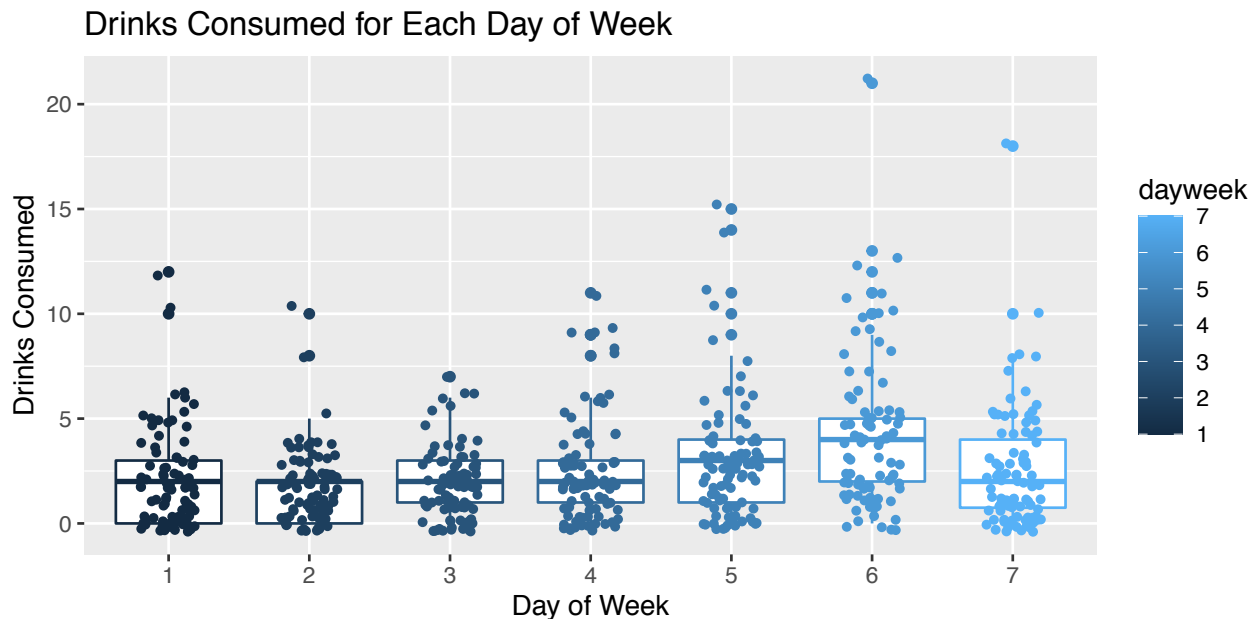
Given the count nature of the *numall* and *desired* variables, we will assume they follow Poisson distributions and use a Poisson modeling approach. A Poisson approach is based on counts, so we must ensure that our dependent variables are counts. The EDA above confirms that *numall* values are all positive integers, but some values for *desired* are repeating decimals that suggest this variable is the result of dividing some total by three. Therefore we will transform *desired* to integer count values by multiplying by 3. Poisson distributions require the count data to have the same underlying process intensity and a constant period of observation for each count. We can satisfy these requirements by limiting our observations to just one day of the week. People's drinking behavior may vary in intensity across a week but we assume that the intensity is the same on one particular day. In addition, the same day of week means the same period of observation.

Now we will evaluate the different days of the week for the dataset using the *dayweek* variable:

```
### Boxplot with dot plot overlay for dayweek vs. numall
plot.2a <- ggplot(q2, aes(x=as.factor(dayweek), y=numall, color=dayweek))
plot.2a <- plot.2a + geom_boxplot() + geom_jitter(shape=16, position=position_jitter(0.2))
plot.2a <- plot.2a + labs(title="Drinks Consumed for Each Day of Week",
                          x="Day of Week", y="Drinks Consumed")
plot.2a
```

```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```

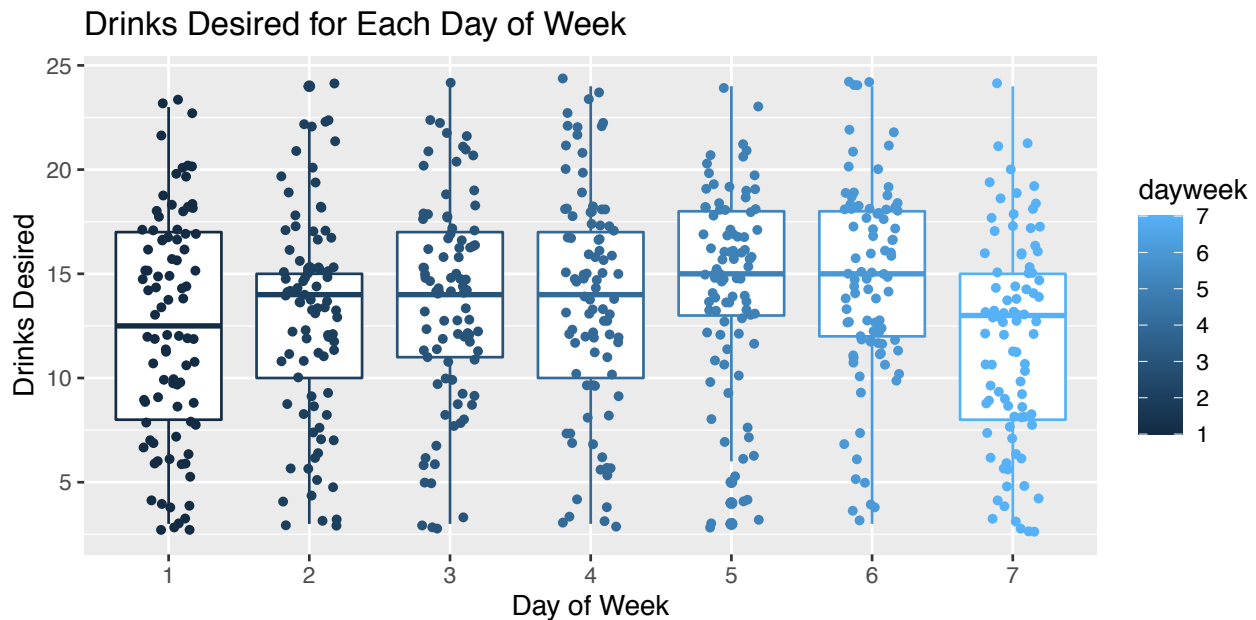
```
## Warning: Removed 1 rows containing missing values (geom_point).
```



```
### Boxplot with dot plot overlay for dayweek vs. desired
plot.2b <- ggplot(q2, aes(x=as.factor(dayweek), y=desired, color=dayweek))
plot.2b <- plot.2b + geom_boxplot() + geom_jitter(shape=16, position=position_jitter(0.2))
plot.2b <- plot.2b + labs(title="Drinks Desired for Each Day of Week",
                          x="Day of Week", y="Drinks Desired")
plot.2b
```

```
## Warning: Removed 3 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 3 rows containing missing values (geom_point).
```



From the day of week plots, Saturday (6) as the widest distribution of count values for *numall* and *desired*, so we will focus on Saturday only for the rest of the analysis.

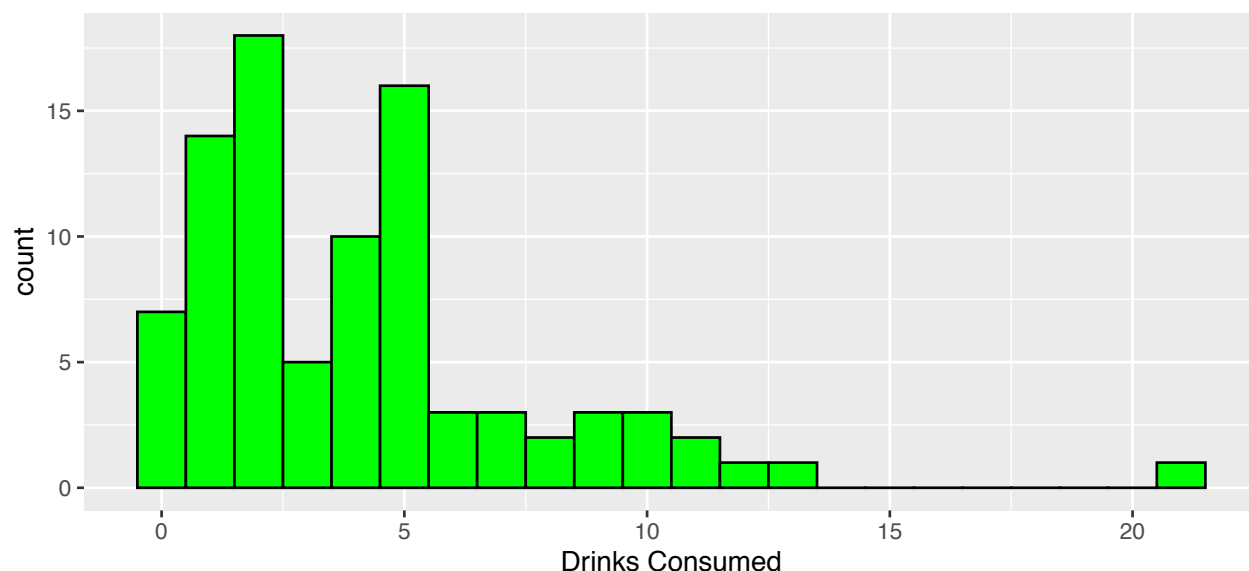
Now we will explore event types, drink consumption and desire. According to the Bilder textbook (page 260), the *negevent* and *posevent* variables are averages of 10 values that include *nrel* and *prel* respectively. Therefore we will create the new variables *nother* and *pother* to separate out the romantic relationship variables to avoid variable overlap.

```
### Filter dataset to include Saturdays only
q2sat <- q2[q2$dayweek==6, ]

### Create nother and pother variables
q2sat$nother <- q2sat$negevent*10 - q2sat$nrel
q2sat$pother <- q2sat$posevent*10 - q2sat$prel

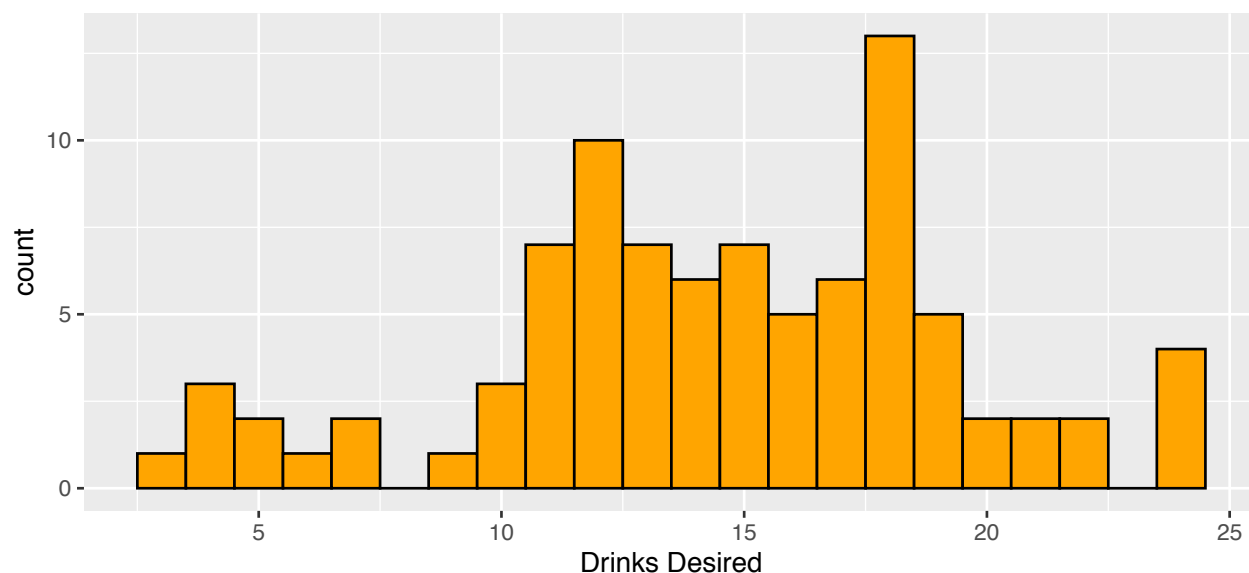
### Histogram for numall
plot.2c <- ggplot(q2sat, aes(x=numall))
plot.2c <- plot.2c + geom_histogram(binwidth=1, color="black", fill="green")
plot.2c <- plot.2c + labs(title="Saturday Drinks Consumed Histogram",
                          x="Drinks Consumed")
plot.2c
```

Saturday Drinks Consumed Histogram



```
### Histogram for desired
plot.2d <- ggplot(q2sat, aes(x=desired))
plot.2d <- plot.2d + geom_histogram(binwidth=1, color="black", fill="orange")
plot.2d <- plot.2d + labs(title="Saturday Drinks Desired Histogram",
                          x="Drinks Desired")
plot.2d
```

Saturday Drinks Desired Histogram



The drinks consumed on Saturdays peak in the 0 to 5 range with a long tail to the right while the drinks desired on Saturdays peak in the 12 to 18 range with tails on both sides. Both of the above plots demonstrate that it is plausible for the dependent variables to follow Poisson distributions.

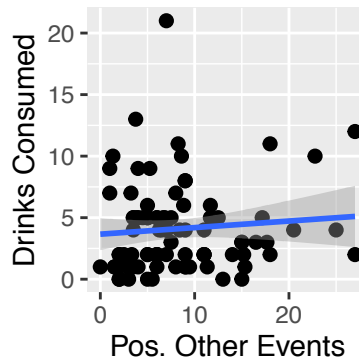
```
### Scatterplot for pother vs. numall
plot.2e <- ggplot(q2sat, aes(x=pother, y=numall))
plot.2e <- plot.2e + geom_point(color="black", size=2)
```

```

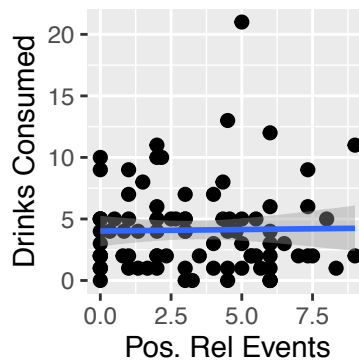
plot.2e <- plot.2e + geom_smooth(method="lm", formula=y~x)
plot.2e <- plot.2e + labs(x="Pos. Other Events", y="Drinks Consumed")

### Scatterplot for prel vs. numall
plot.2f <- ggplot(q2sat, aes(x=prel, y=numall))
plot.2f <- plot.2f + geom_point(color="black", size=2)
plot.2f <- plot.2f + geom_smooth(method="lm", formula=y~x)
plot.2f <- plot.2f + labs(x="Pos. Rel Events", y="Drinks Consumed")
plot.2e

```



```
plot.2f
```



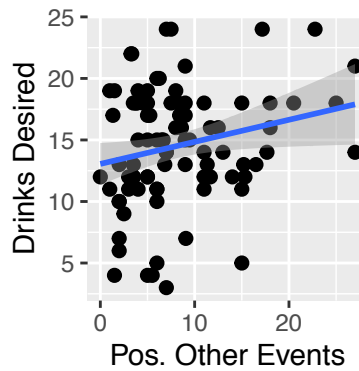
The above scatterplots indicate that positive events of any type generally have no relationship with amount of drinks consumed.

```

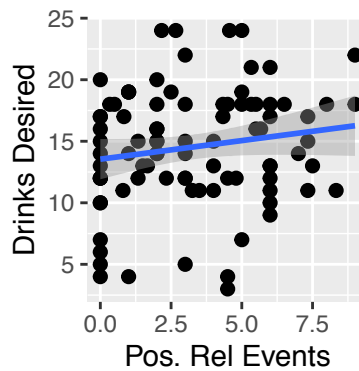
### Scatterplot for pother vs. desired
plot.2g <- ggplot(q2sat, aes(x=pother, y=desired))
plot.2g <- plot.2g + geom_point(color="black", size=2)
plot.2g <- plot.2g + geom_smooth(method="lm", formula=y~x)
plot.2g <- plot.2g + labs(x="Pos. Other Events", y="Drinks Desired")

### Scatterplot for prel vs. desired
plot.2h <- ggplot(q2sat, aes(x=prel, y=desired))
plot.2h <- plot.2h + geom_point(color="black", size=2)
plot.2h <- plot.2h + geom_smooth(method="lm", formula=y~x)
plot.2h <- plot.2h + labs(x="Pos. Rel Events", y="Drinks Desired")
plot.2g

```

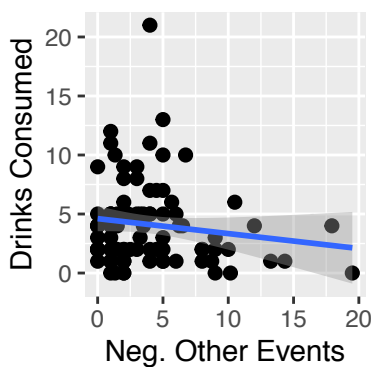
```
plot.2h
```



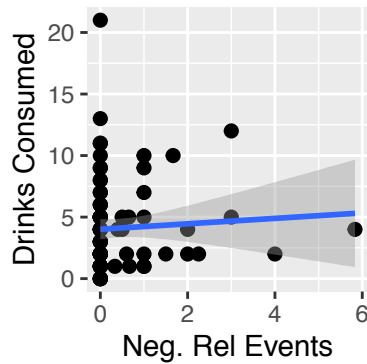
The above scatterplots indicate that positive events of any type have a weakly positive relationship with drinks desired.

```
### Scatterplot for nother vs. numall
plot.2i <- ggplot(q2sat, aes(x=nother, y=numall))
plot.2i <- plot.2i + geom_point(color="black", size=2)
plot.2i <- plot.2i + geom_smooth(method="lm", formula=y~x)
plot.2i <- plot.2i + labs(x="Neg. Other Events", y="Drinks Consumed")

### Scatterplot for nrel vs. numall
plot.2j <- ggplot(q2sat, aes(x=nrel, y=numall))
plot.2j <- plot.2j + geom_point(color="black", size=2)
plot.2j <- plot.2j + geom_smooth(method="lm", formula=y~x)
plot.2j <- plot.2j + labs(x="Neg. Rel Events", y="Drinks Consumed")
plot.2i
```



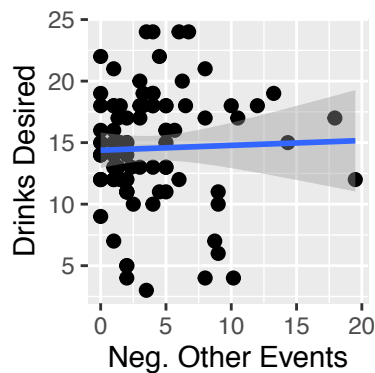
```
plot.2j
```



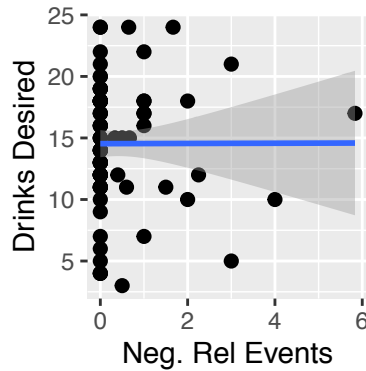
The above scatterplots indicate some weak negative relationship between negative other events and drinks consumed. However, no clear relationship is seen between negative relationship events and drinks consumed.

```
### Scatterplot for nother vs. desired
plot.2k <- ggplot(q2sat, aes(x=nother, y=desired))
plot.2k <- plot.2k + geom_point(color="black", size=2)
plot.2k <- plot.2k + geom_smooth(method="lm", formula=y~x)
plot.2k <- plot.2k + labs(x="Neg. Other Events", y="Drinks Desired")

### Scatterplot for nrel vs. desired
plot.2l <- ggplot(q2sat, aes(x=nrel, y=desired))
plot.2l <- plot.2l + geom_point(color="black", size=2)
plot.2l <- plot.2l + geom_smooth(method="lm", formula=y~x)
plot.2l <- plot.2l + labs(x="Neg. Rel Events", y="Drinks Desired")
plot.2k
```



```
plot.2l
```



The above scatterplots indicate that negative events of any type generally have no relationship with amount of drinks desired.

In totality, the graphical evidence suggests that negative romantic interactions have no associations with alcohol consumption and desire to drink, contradicting the original researcher hypotheses.

Now we will investigate trait self-esteem in the Saturday cohort:

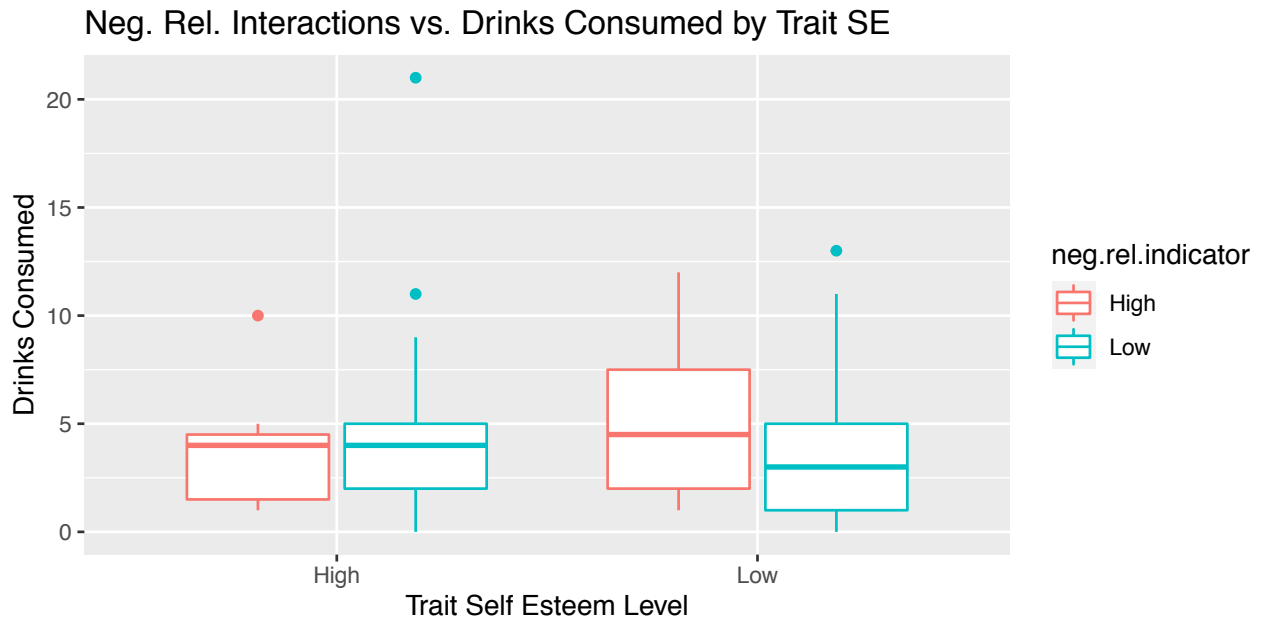
```
### Add variable to indicate low and high trait self-esteem groups
### Low is <= median, High is > median
rosn.median <- median(q2sat$rosn)
rosn.ind <- function(x) {ifelse(x<=rosn.median, yes="Low", no="High")}
q2sat$trait.self.esteem <- as.factor(sapply(q2sat$rosn, FUN=rosn.ind))

### Add indicator variable for amount of negative relationship interactions
### Low is <= median, High is > median
nrel.median <- median(q2sat$nrel)
nrel.ind <- function(y) {ifelse(y<=nrel.median, yes="Low", no="High")}
q2sat$neg.rel.indicator <- as.factor(sapply(q2sat$nrel, FUN=nrel.ind))

### Trivariate plot: trait self-esteem, neg. rel. interactions and drinks consumed
plot.2m <- ggplot(q2sat, aes(x=trait.self.esteem, y=numall,
                             color=neg.rel.indicator))
plot.2m <- plot.2m + geom_boxplot()
plot.2m <- plot.2m + labs(title="Neg. Rel. Interactions vs. Drinks Consumed by Trait SE",
                          x="Trait Self Esteem Level", y="Drinks Consumed")

### Trivariate plot: trait self-esteem, neg. rel. interactions and drinks desired
plot.2n <- ggplot(q2sat, aes(x=trait.self.esteem, y=desired,
                             color=neg.rel.indicator))
plot.2n <- plot.2n + geom_boxplot()
plot.2n <- plot.2n + labs(title="Neg. Rel. Interactions vs. Drinks Desired by Trait SE",
                          x="Trait Self Esteem Level", y="Drinks Desired")
```

plot.2m



plot.2n

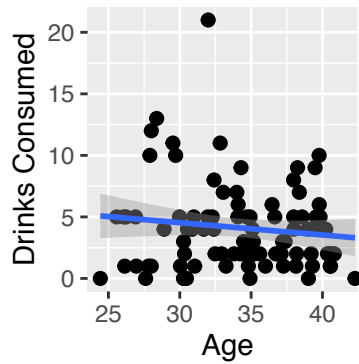


Regarding the hypothesis on people with low trait self-esteem, the evidence of more drinking for more negative relationship interactions (“NRIs”) from the above boxplots is marginal at best. In terms of drinks desired, the difference between days of more or less NRIs is negligible for low trait self-esteem people. In terms of drinks consumed, while the median level is slightly higher for days with high NRIs, there is enough overlap in values between high/low NRIs that the difference may not be statistically or practically significant.

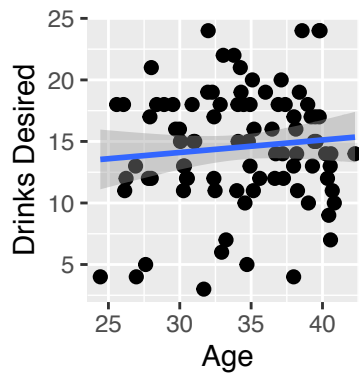
Researchers hypothesized the lack of relationship between drinking and NRIs for people with high trait self-esteem. The boxplots confirm this view with negligible differences for both drinks consumed and drinks desired among different NRI levels for high trait self-esteem people.

Finally, we will examine age and state self-esteem relative to drinking patterns:

```
### Scatterplots for age vs. drinks consumed/desired
plot.2o <- ggplot(q2sat, aes(x=age, y=numall))
plot.2o <- plot.2o + geom_point(size=2)
plot.2o <- plot.2o + geom_smooth(method="lm", formula=y~x)
plot.2o <- plot.2o + labs(x="Age", y="Drinks Consumed")
plot.2o
```

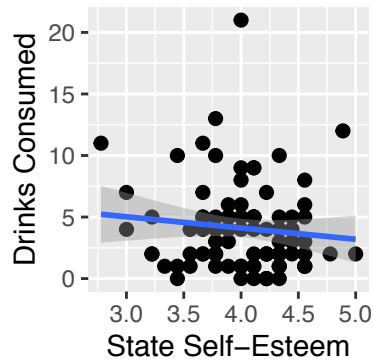


```
plot.2p <- ggplot(q2sat, aes(x=age, y=desired))
plot.2p <- plot.2p + geom_point(size=2)
plot.2p <- plot.2p + geom_smooth(method="lm", formula=y~x)
plot.2p <- plot.2p + labs(x="Age", y="Drinks Desired")
plot.2p
```

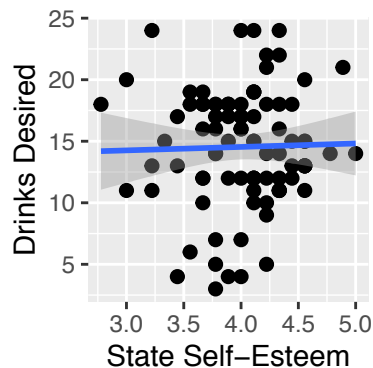


The plots on age indicate contrasting relationships. While age has a weakly negative relationship with drinks consumed, it has a weakly positive relationship with drinks desired.

```
### Scatterplots for state self-esteem vs. drinks consumed/desired
plot.2q <- ggplot(q2sat, aes(x=state, y=numall))
plot.2q <- plot.2q + geom_point(size=2)
plot.2q <- plot.2q + geom_smooth(method="lm", formula=y~x)
plot.2q <- plot.2q + labs(x="State Self-Esteem", y="Drinks Consumed")
plot.2q
```



```
plot.2r <- ggplot(q2sat, aes(x=state, y=desired))
plot.2r <- plot.2r + geom_point(size=2)
plot.2r <- plot.2r + geom_smooth(method="lm", formula=y~x)
plot.2r <- plot.2r + labs(x="State Self-Esteem", y="Drinks Desired")
plot.2r
```



From a state self-esteem perspective, there is only a slight negative relationship with drinks consumed. Any relationship with drinks desired looks non-existent.

Question 2.2: Negative Romantic Interactions Hypothesis

Researchers originally hypothesized that negative relationship interactions (“NRIs”) are associated with higher drinks consumed and drinks desired. We will investigate this by estimating Poisson regression models.

For drinks consumed, our exploratory data analysis found that the *nother*, *age*, *rosn*, and *state* variables had at least some weak relationship with *numall*. Therefore we will include these explanatory variables in our model in addition to our main explanatory variable of interest, *nrel*.

```
### Estimate drinks consumed Poisson model
q2.model1 <- glm(numall ~ nrel + nother + age + rosn + state,
                 family = poisson(link="log"), data = q2sat)
summary(q2.model1)

##
## Call:
## glm(formula = numall ~ nrel + nother + age + rosn + state, family = poisson(link = "log"),
##      data = q2sat)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.1987  -1.3498  -0.3857   0.5210   5.9009
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.58361    0.72346   4.953 7.29e-07 ***
## nrel           0.10315    0.05478   1.883  0.05968 .
## nother        -0.05124    0.01611  -3.180  0.00147 **
## age           -0.02693    0.01172  -2.298  0.02159 *
## rosn           0.02589    0.13060   0.198  0.84284
## state        -0.29992    0.12666  -2.368  0.01788 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 250.34  on 88  degrees of freedom
## Residual deviance: 230.79  on 83  degrees of freedom
## AIC: 498.19
##
## Number of Fisher Scoring iterations: 5
```

Our drinks consumed model can be expressed as the following:

$$Y_i \sim Po(\mu_i)$$

$$\text{with } \log(\mu_i) = 3.5836 + 0.1032nrel - 0.0512nother - 0.0269age + 0.0259rosn - 0.2999state$$

where Y_i is number of drinks consumed for person i

Now we will conduct a LRT to test the *nrel* variable at a 5% level of significance and construct a 95% confidence interval:

```
### LRT for the drinks consumed model
Anova(q2.model1)
```

```
## Analysis of Deviance Table (Type II tests)
##
```

```
## Response: numall
##      LR Chisq Df Pr(>Chisq)
## nrel      3.2698 1 0.0705663 .
## nother    11.1076 1 0.0008597 ***
## age       5.2502 1 0.0219441 *
## rosn      0.0394 1 0.8426316
## state     5.5457 1 0.0185259 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

### 95% confidence interval for nrel converted to percentage change
100*(exp(confint(q2.model1, parm="nrel", level=0.95)) - 1)
```

```
## Waiting for profiling to be done...
```

```
##      2.5 %      97.5 %
## -0.9009071 22.9025783
```

nrel has a p-value of 0.07 and its percentage change confidence interval includes zero. Thus, there is not enough evidence that *nrel* has an effect on drinks consumed.

For drinks desired, our exploratory data analysis found that the *prel*, *pother*, and *age* variables had at least some weak relationship with *desired*. Therefore we will include these explanatory variables in our model in addition to our main explanatory variable of interest, *nrel*.

```
### Estimate drinks desired Poisson model
q2.model2 <- glm(desired ~ nrel + prel + pother + age,
                 family = poisson(link="log"), data = q2sat)
summary(q2.model2)
```

```
##
## Call:
## glm(formula = desired ~ nrel + prel + pother + age, family = poisson(link = "log"),
##      data = q2sat)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6541  -0.8241   0.0106   0.7724   2.2757
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.340525   0.220300  10.624  <2e-16 ***
## nrel         0.002091   0.030566   0.068  0.9455
## prel        0.016447   0.011539   1.425  0.1541
## pother      0.009829   0.004692   2.095  0.0362 *
## age         0.005721   0.006271   0.912  0.3616
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 153.89  on 88  degrees of freedom
## Residual deviance: 144.36  on 84  degrees of freedom
## AIC: 550.94
##
## Number of Fisher Scoring iterations: 4
```


Our drinks desired model can be expressed as the following:

$$Y_i \sim Po(\mu_i)$$

$$\text{with } \log(\mu_i) = 2.3405 + 0.0021nrel + 0.0164prel + 0.0098pother + 0.0057age$$

where Y_i is number of drinks desired for person i

Now we will conduct a LRT to test the *nrel* variable at a 5% level of significance and construct a 95% confidence interval:

```
### LRT for the drinks desired model
Anova(q2.model12)

## Analysis of Deviance Table (Type II tests)
##
## Response: desired
##      LR Chisq Df Pr(>Chisq)
## nrel    0.0047  1    0.94552
## prel    2.0255  1    0.15468
## pother   4.3110  1    0.03787 *
## age     0.8346  1    0.36096
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

### 95% confidence interval for nrel converted to percentage change
100*(exp(confint(q2.model12, parm="nrel", level=0.95)) - 1)

## Waiting for profiling to be done...

##      2.5 %      97.5 %
## -5.780113  6.221904
```

nrel has a p-value of 0.95 and its percentage change confidence interval includes zero. Thus, there is not enough evidence that *nrel* has an effect on drinks desired.

Although our two Poisson models for drinks consumed and drinks desired had positive coefficients for NRIs, neither was statistically significant given the presence of other variables in the model. Thus we do not have enough evidence for the researcher hypothesis that NRIs are associated with higher drinks consumed and desired. Instead, other variables such as other negative events, other positive events, age and state self-esteem have stronger evidence for explaining drinking behavior.

Question 2.3: High Trait Self-Esteem Individuals

Researchers originally hypothesized that the relationship between drinking behavior and negative relationship interactions (“NRIs”) should not be evident for people with high trait self-esteem. We will investigate this by estimating Poisson regression models for the high trait self-esteem population with the same variables as those used in Question 2.2.

```
### Filter Saturday dataset for high trait self-esteem
q2sat.hi.trait <- q2sat[q2sat$trait.self.estesteem=="High", ]

### Estimate drinks consumed Poisson model
q2.model13 <- glm(numall ~ nrel + nother + age + rosn + state,
                  family = poisson(link="log"), data = q2sat.hi.trait)
summary(q2.model13)

##
## Call:
```

```
## glm(formula = numall ~ nrel + nother + age + rosn + state, family = poisson(link = "log"),
##      data = q2sat.hi.trait)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3414  -1.4315  -0.1462   0.5509   5.0402
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  7.55061    2.34330   3.222  0.00127 **
## nrel          0.03818    0.08323   0.459  0.64638
## nother       -0.05410    0.02192  -2.468  0.01359 *
## age          -0.00979    0.02002  -0.489  0.62491
## rosn         -1.19213    0.56595  -2.106  0.03517 *
## state        -0.25722    0.20597  -1.249  0.21173
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 112.949  on 39  degrees of freedom
## Residual deviance:  98.652  on 34  degrees of freedom
## AIC: 228.79
##
## Number of Fisher Scoring iterations: 5
```

Our drinks consumed model for high trait self-esteem individuals can be expressed as the following:

$$Y_i \sim Po(\mu_i)$$

with $\log(\mu_i) = 7.5506 + 0.0382nrel - 0.0541nother - 0.0098age - 1.1921rosn - 0.2572state$
 where Y_i is number of drinks consumed for person i

Now we will conduct a LRT to test the *nrel* variable at a 5% level of significance and construct a 95% confidence interval:

```
### LRT for the high trait self-esteem drinks consumed model
Anova(q2.model3)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: numall
##      LR Chisq Df Pr(>Chisq)
## nrel    0.2053  1  0.650442
## nother   6.7808  1  0.009214 **
## age      0.2387  1  0.625146
## rosn     4.4903  1  0.034088 *
## state    1.5572  1  0.212078
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
### 95% confidence interval for nrel converted to percentage change
100*(exp(confint(q2.model3, parm="nrel", level=0.95)) - 1)
```

```
## Waiting for profiling to be done...
```

```
##      2.5 %      97.5 %
```

```
## -12.57399 21.35513
```

nrel has a p-value of 0.65 and its percentage change confidence interval includes zero. Thus, there is not enough evidence that *nrel* has an effect on drinks consumed.

Next, we will examine drinks desired for high trait self-esteem individuals:

```
### Estimate drinks desired Poisson model
```

```
q2.model4 <- glm(desired ~ nrel + prel + pother + age,  
                 family = poisson(link="log"), data = q2sat.hi.trait)  
summary(q2.model4)
```

```
##  
## Call:  
## glm(formula = desired ~ nrel + prel + pother + age, family = poisson(link = "log"),  
##      data = q2sat.hi.trait)  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max  
## -3.4971  -0.6262  -0.0841   0.6949   2.4899  
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)  
## (Intercept)  2.433923   0.372646   6.531 6.51e-11 ***  
## nrel         -0.020252   0.040399  -0.501 0.616158  
## prel         -0.004782   0.017162  -0.279 0.780523  
## pother        0.024184   0.006924   3.493 0.000478 ***  
## age           0.001361   0.011111   0.123 0.902484  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for poisson family taken to be 1)  
##  
##      Null deviance: 63.916  on 39  degrees of freedom  
## Residual deviance: 51.609  on 35  degrees of freedom  
## AIC: 240.05  
##  
## Number of Fisher Scoring iterations: 4
```

Our drinks desired model for those with high trait self-esteem can be expressed as the following:

$$Y_i \sim Po(\mu_i)$$

$$\text{with } \log(\mu_i) = 2.4339 - 0.0203nrel - 0.0048prel + 0.0242pother + 0.0014age$$

where Y_i is number of drinks desired for person i

Now we will conduct a LRT to test the *nrel* variable at a 5% level of significance and construct a 95% confidence interval:

```
### LRT for the high trait self-esteem drinks desired model
```

```
Anova(q2.model4)
```

```
## Analysis of Deviance Table (Type II tests)  
##  
## Response: desired  
##      LR Chisq Df Pr(>Chisq)  
## nrel    0.2562  1  0.6127724
```

```
## prel      0.0777  1  0.7804485
## pother    11.7567  1  0.0006062 ***
## age       0.0150  1  0.9024747
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

### 95% confidence interval for nrel converted to percentage change
100*(exp(confint(q2.model4, parm="nrel", level=0.95)) - 1)

## Waiting for profiling to be done...
##      2.5 %      97.5 %
## -9.742060  5.769802
```

nrel has a p-value of 0.61 and its percentage change confidence interval includes zero. Thus, there is not enough evidence that *nrel* has an effect on drinks desired.

In totality, neither of our models found *nrel* to be statistically significant given the presence of other variables in the model. Thus, we reach a conclusion consistent with the original researcher hypothesis. Among individuals with high trait self-esteem, there is no evidence of a relationship between negative relationship interactions and greater drink consumption and desire.