Lab 1, Short Questions

Contents

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.*

```r
cereal <- read_csv("../data/short-questions/cereal_dillons.csv")

# Rename raw data columns to keep final names short
names(cereal) <- c("ID", "Shelf", "Cereal", "size_g",
                   "sugar_raw", "fat_raw", "sodium_raw")
```

1.1 Recode Data

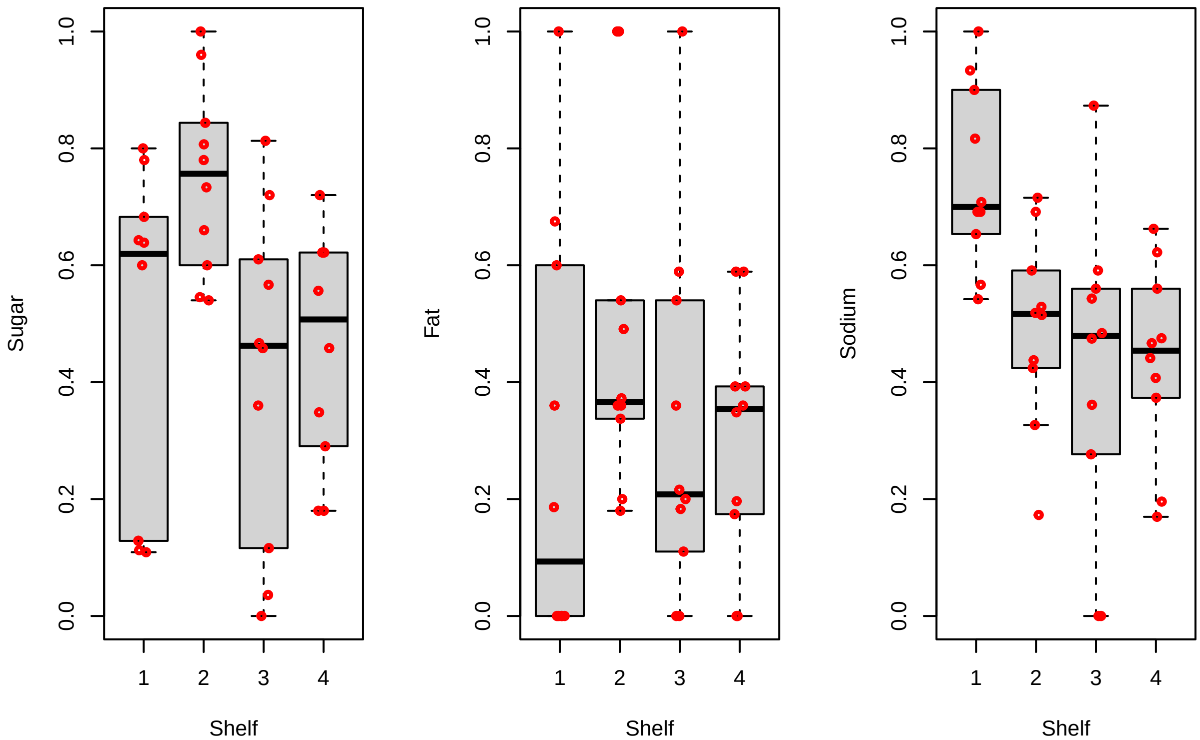
(1 point) The explanatory variables need to be reformatted before proceeding further (sample code is provided in the textbook). First, divide each explanatory variable by its serving size to account for the different serving sizes among the cereals. Second, rescale each variable to be within 0 and 1. 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.

```r
# Re-scale columns in the following list
# Create new columns for re-scaled data
col_trans_list <- c("fat_raw", "sugar_raw", "sodium_raw")

for (col in col_trans_list) {
  col_per <- cereal[col] / cereal["size_g"]
  cereal[str_sub(col, 1, -5)] <- (col_per - min(col_per)) / max(col_per)
}

cereal
```

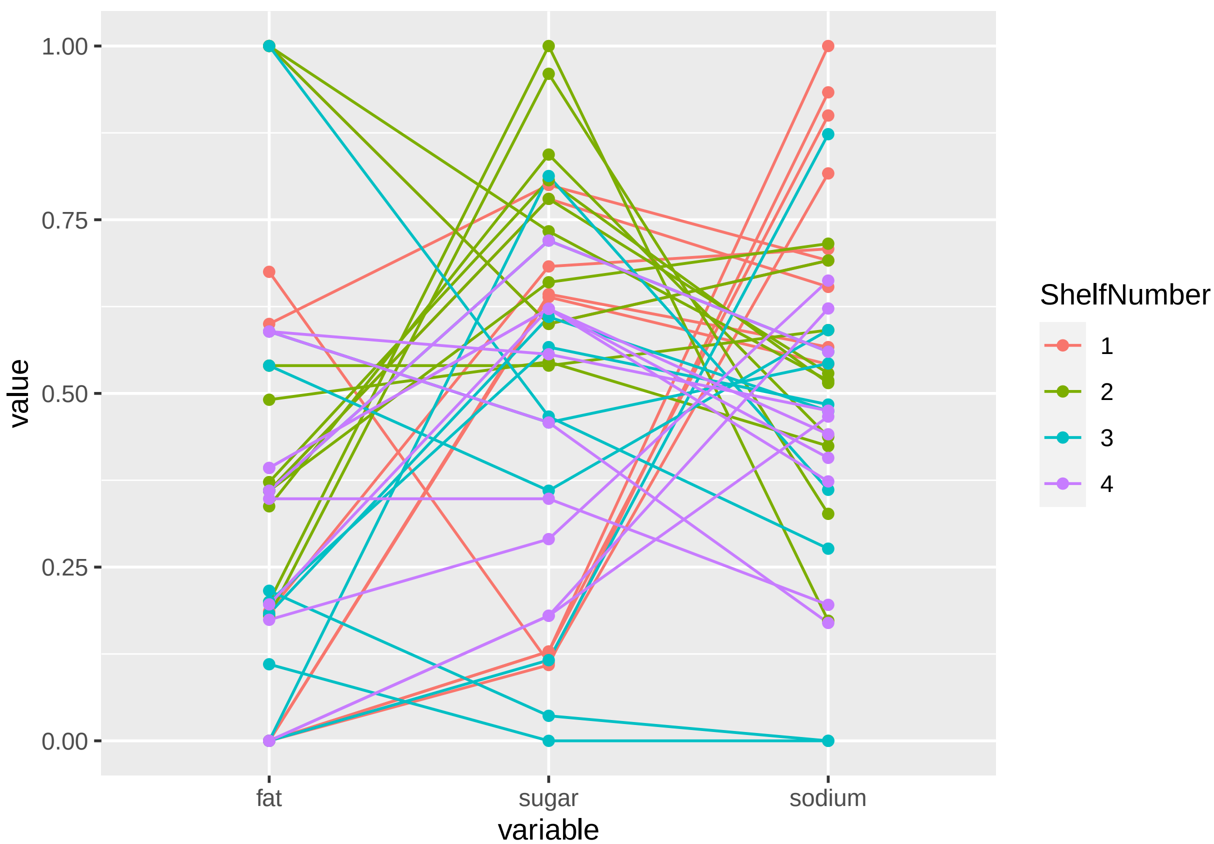
```
## # A tibble: 40 x 10
##       ID Shelf Cereal     size_g sugar_raw fat_raw sodium_raw   fat sugar sodium
##    <dbl> <dbl> <chr>       <dbl>     <dbl>   <dbl>      <dbl> <dbl> <dbl>  <dbl>
##  1     1     1 Kellog's ~     28        10     0          170 0     0.643  0.567
##  2     2     1 Post Toas~     28         2     0          270 0     0.129  0.9
##  3     3     1 Kellogg's~     28         2     0          300 0     0.129  1
##  4     4     1 Food Club~     32         2     2          280 0.675 0.112  0.817
##  5     5     1 Frosted C~     30        13     1          210 0.36  0.78   0.653
##  6     6     1 Food Club~     31        11     0          180 0     0.639  0.542
##  7     7     1 Capn Crun~     27        12     1.5        200 0.6   0.8    0.691
##  8     8     1 Capn Crun~     27         9     2.5        200 1     0.6    0.691
##  9     9     1 Post Hone~     29        11     0.5        220 0.186 0.683  0.708
## 10    10     1 Food Club~     33         2     0          330 0     0.109  0.933
## # i 30 more rows
```

```r
par(mfrow = c(1, 3))

# Sugar
boxplot(formula = sugar ~ Shelf, data = cereal, ylab = "Sugar",
        xlab = "Shelf", pars = list(outpch = NA))
stripchart(x = cereal$sugar ~ cereal$Shelf, lwd = 2,
           col = "red", method = "jitter", vertical = TRUE, pch = 1, add = TRUE)

# Fat
boxplot(formula = fat ~ Shelf, data = cereal, ylab = "Fat",
        xlab = "Shelf", pars = list(outpch = NA))
stripchart(x = cereal$fat ~ cereal$Shelf, lwd = 2,
           col = "red", method = "jitter", vertical = TRUE, pch = 1, add = TRUE)

# Sodium
boxplot(formula = sodium ~ Shelf, data = cereal, ylab = "Sodium",
        xlab = "Shelf", pars = list(outpch = NA))
stripchart(x = cereal$sodium ~ cereal$Shelf, lwd = 2,
           col = "red", method = "jitter", vertical = TRUE, pch = 1, add = TRUE)
```



It appears that the items that are higher in sodium tend to be on shelf #1. Items that are high in sugar tend to go on shelf #2. For fat, there does not appear to be any obvious trends of which shelf items that are high or low in fat go to.

```
cereal$ShelfNumber <- as.character(cereal$Shelf)
ggparcoord(data = cereal, columns = 8:10,groupColumn = "ShelfNumber",
           showPoints = TRUE, scale = "uniminmax")
```



The parallel coordinates plot is harder to glean information from than the box plots. When looking closely one can find much of the same information that was found in the box plots. Namely, the highest sugar items belonging to shelf #2 and the high sodium items being found in shelf #1.

Without running a regression analysis it is not responsible to say definitively whether or not a content different exists between the shelves. Based on these graphs there doesn't seem to be too big of a relationship between sugar, fat, and sodium and the shelf a cereal belongs to.

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).

Ordinal data would make sense in the case where there is a natural ordering to the shelves. Such as if higher shelves were inherently more desirable than lower shelves. Then it would be expected that better selling products be placed on higher shelves. It does not appear as if this is the case here as the most desirable shelves would likely be at eye-level (towards the middle) and there doesn't appear to be any other sort of ordering at play.

```r
# Set Shelf as a categorical value
cereal$Shelf <- factor(cereal$Shelf, levels = c("1", "2", "3", "4"))

# Estimate linear model
model_cereal_shelves_linear <- multinom(
  formula = Shelf ~ fat + sugar + sodium,
  data = cereal
)

# Estimate linear model with all interactions
model_cereal_shelves_quadratic <- multinom(
  formula = Shelf ~ fat + sugar + sodium +
    fat:sugar + fat:sodium + sodium:sugar +
    sodium:sugar:fat,
  data = cereal
)
```

```r
# Conduct Anova test on linear model
lrt_cereal_main_effects <- car::Anova(model_cereal_shelves_linear, test = "LR")
lrt_cereal_main_effects
```

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

```r
# Conduct Anova test on the interactions
lrt_cereal_quadratic_effects <- car::Anova(model_cereal_shelves_quadratic, test = "LR")
lrt_cereal_quadratic_effects
```

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

The LRT test for linear response variables shows strong statistical significance for sugar and sodium content (p-values of 4.520699×10^{-5} and 7.0732813×10^{-6} respectievely) but fails to show even weak significance of fat content (p-value of 0.1522)

The LRT test for the interactions further revealed that no two-way or three way interaction achieved statistical significance

Similar to what the graphs show there is clearly an obvious relationship between sugar, sodium, and shelf placement. The Anova test shows that none of the interaction terms are significant as well as the linear fat variable.

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.

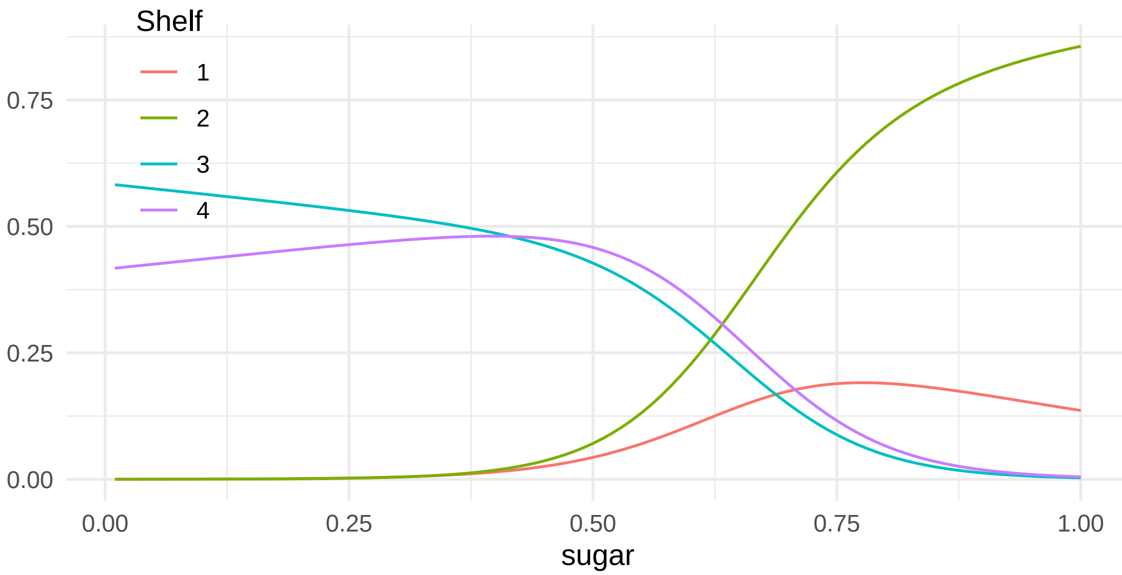
```
# Estimate new model that removes non-significant fat variable
model_cereal_shelves_trim <- multinom(formula = Shelf ~ sugar + sodium,
                                      data = cereal)

# Create a datframe with Apple Jack data
app_jack <- data.frame(size_g = 28,
                       sugar_raw = 12,
                       fat_raw = 0.5, sodium_raw = 130)

# Use the same normalization procedures as for the main dataframe
for (col in col_trans_list) {
  # Column of variable devided by portion size:
  col_per <- app_jack[col] / app_jack["size_g"]

  # Reference column of variable devided by portion size:
  ref_col <- cereal[col] / cereal["size_g"]
  app_jack[str_sub(col, 1, -5)] <- (col_per - min(ref_col)) / max(ref_col)
}

# Estimate placement of Apple Jack with a trimmed model
aj_shelf_probs_trim <- predict(model_cereal_shelves_trim,
                               newdata = app_jack, type = "probs")
shelf_trim <- aj_shelf_probs_trim[which.max(aj_shelf_probs_trim)]

# Estimate placement of Apple Jack with bloated model
aj_shelf_probs <- predict(model_cereal_shelves_linear,
                          newdata = app_jack, type = "probs")
shelf <- aj_shelf_probs[which.max(aj_shelf_probs)]

aj_shelf_probs
```

Using best practices for variable selection we estimated a new model that only contains statistically significant variables: sugar and sodium. Using this model we estimate probabilities of placing Apple Jack on the shelves 1, 2, 3 and 4 respectively as 4, 59, 16, 21 percentage points. Thus, shelf 2 is clearly the most likely place. If we were to use a full model, that includes insignificant variable Fat, the result would stay the same, but the difference between shelves would be smaller at 5, 47, 20, 27 percentage points.

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.

```r
# Get mean values for static values
fat_mean <- mean(cereal$fat)
na_mean <- mean(cereal$sodium)

# Make dataframe with values used fo inference
df_to_plot <- data.frame(fat = rep(fat_mean, times = 100),
                         sodium = rep(na_mean, times = 100),
                         sugar = seq(1, 100) / 100)

# Attach predicted values to this dataframe
df_to_plot <- cbind(df_to_plot,
    predict(model_cereal_shelves_linear, newdata = df_to_plot, type = "probs"))

# Plot the data
shelf_vs_sugar_plot <- df_to_plot %>%
  pivot_longer(cols = c(4:7), names_to = "indicators", values_to = "values") %>%
  ggplot(data = ., aes(x = sugar, y = values, colour = indicators)) +
  geom_line() +
  theme_minimal() +
  theme(legend.position = c(0.1, 0.8)) + guides(color=guide_legend("Shelf")) +
  theme(axis.title.y = element_blank())

shelf_vs_sugar_plot
```



For cereals with normalized sugar content up to approximately average, there is roughly eqaul chance of finding them on 4'th and 3'd shelfs. Assuming the first shelf is at teh bottom and the 4th is at the top, an average health-concious adult might see them there. As the normalized sugar content approaches the higher end of the spectrum, the chances of finding this cereal on the second shelf, where a kid might see it, is growing dramatically.

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.

```
coefs.2 <- coef(model_cereal_shelves_linear)[1,]
se.2 <- round(summary(model_cereal_shelves_linear)$standard.errors[1,], 2)
ci.2 <- round(data.frame(estimate = exp(coefs.2),
                         lower = exp(coefs.2 - 1.96*se.2),
                         upper =  exp(coefs.2 + 1.96*se.2)), 2)
ci.2
```

```
##             estimate lower        upper
## (Intercept)   992.98  0.00 332153489.14
## fat            58.25  0.63      5390.05
## sugar          14.78  0.00    293922.34
## sodium          0.00  0.00         0.03
```

```
coefs.3 <- coef(model_cereal_shelves_linear)[2,]
se.3 <- round(summary(model_cereal_shelves_linear)$standard.errors[2,], 2)
ci.3 <- round(data.frame(estimate = exp(coefs.3),
                         lower = exp(coefs.3 - 1.96*se.3),
                         upper =  exp(coefs.3 + 1.96*se.3)), 2)
ci.3
```

```
##                 estimate   lower        upper
## (Intercept) 2.604952e+09 1186.40 5.719627e+15
```

```
## fat         5.700000e-01    0.01 6.449000e+01
## sugar       0.000000e+00    0.00 7.000000e-02
## sodium      0.000000e+00    0.00 0.000000e+00
```

```
coefs.4 <- coef(model_cereal_shelves_linear)[3,]
se.4 <- round(summary(model_cereal_shelves_linear)$standard.errors[3,], 2)
ci.4 <- round(data.frame(estimate = exp(coefs.4),
                         lower = exp(coefs.4 - 1.96*se.4),
                         upper =  exp(coefs.4 + 1.96*se.4)), 2)
ci.4
```

```
##                 estimate  lower        upper
## (Intercept) 1.759584e+09 817.25 3.788488e+15
## fat         4.200000e-01   0.00 4.716000e+01
## sugar       0.000000e+00   0.00 1.600000e-01
## sodium      0.000000e+00   0.00 0.000000e+00
```

```
odds_ratios <- c(ci.2, ci.3, ci.4)
```

> One can see that the estimate and corresponding interval for shelf 2's parameters indicate that the odds increase substantially as `fat` and `sugar` increase. It is also seen that the odds increase very slowly for shelves 2-4 as `sodium` increases showing that shelf 1 is the most probable shelf when sodium is high. Lastly, the confidence intervals are fairly wide which is a result of the weak correlation between the variables and shelf number as well as the limited number of samples.

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.*

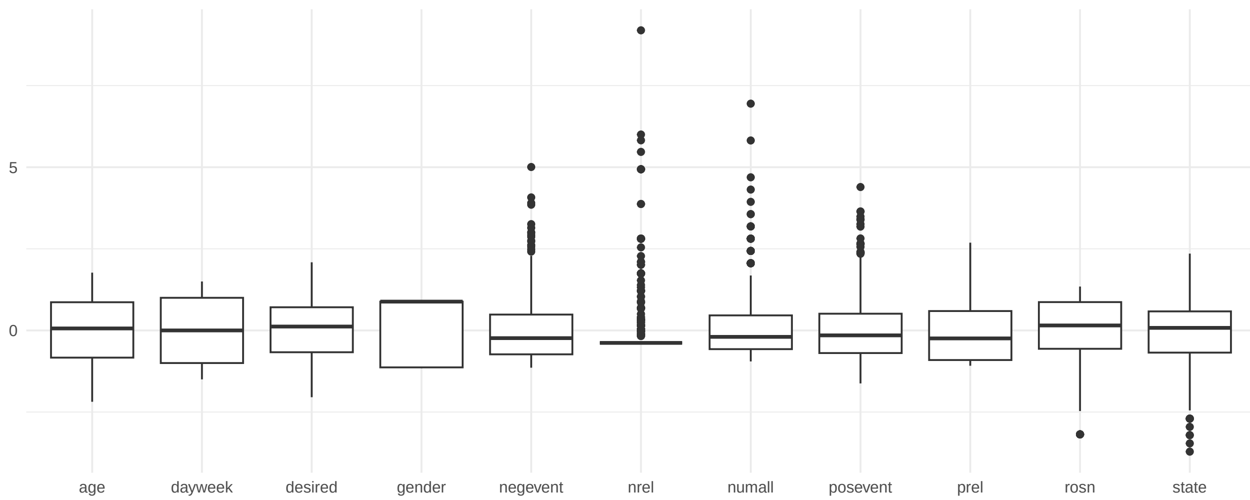
```
drinks <- read_csv('../data/short-questions/DeHartSimplified.csv')
drinks.selected <- select(drinks, "numall", "prel", "nrel",
```

```
                               "age", "rosn", "posevent", "negevent", "desired")
saturday <- filter(drinks, dayweek == "6")
```

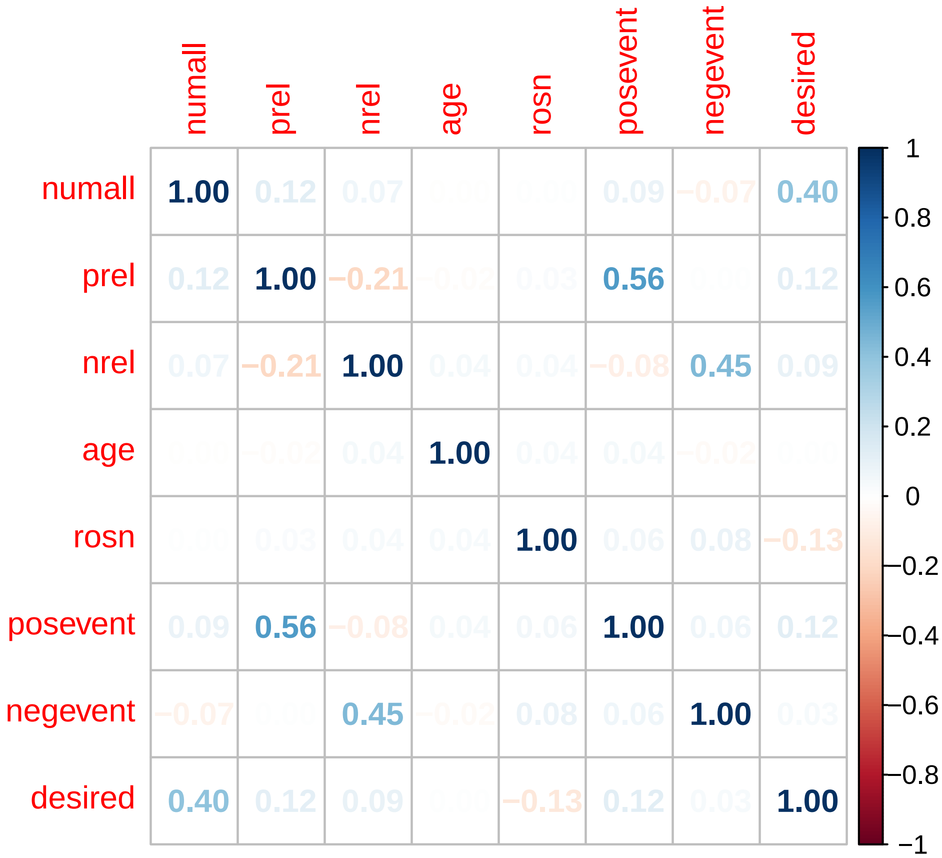
2.1 EDA

(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.

```
drinks_scaled <- as.data.frame(scale(drinks))
drinks_scaled %>%
  pivot_longer(cols = 3:13, names_to = "indicators", values_to = "values") %>%
  ggplot(data = ., aes(x = indicators, y = values)) +
  geom_boxplot() +
  theme_minimal() +
  theme(axis.title.y = element_blank(), axis.title.x = element_blank())
```




```
M = cor(drinks.selected,  use = "pairwise.complete.obs")
corrplot(M, method = 'number')
```

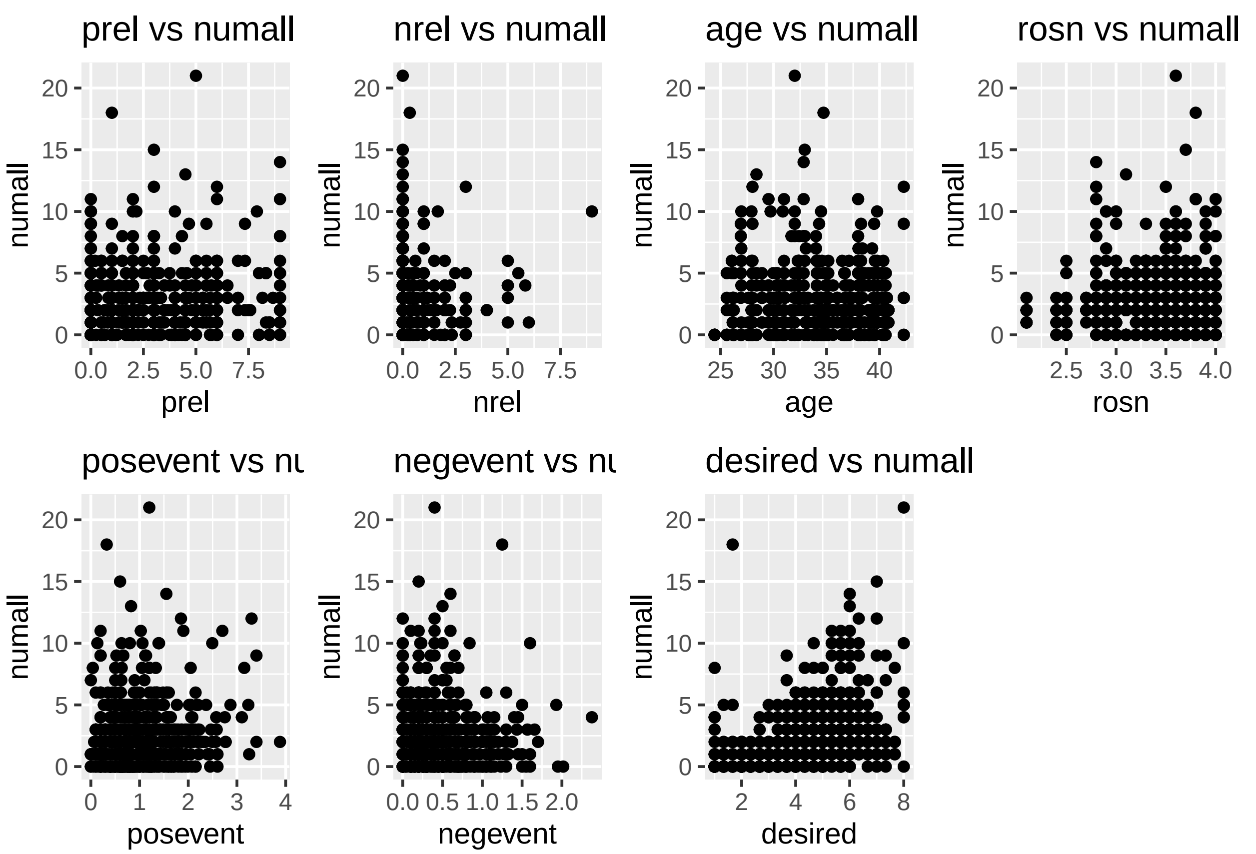



```r
summary(drinks)
```

```
##        id           studyday    dayweek        numall            nrel
##  Min.   :  1.00   Min.   :1   Min.   :1   Min.   : 0.000   Min.   :0.000
##  1st Qu.: 33.00   1st Qu.:2   1st Qu.:2   1st Qu.: 1.000   1st Qu.:0.000
##  Median : 60.00   Median :4   Median :4   Median : 2.000   Median :0.000
##  Mean   : 75.89   Mean   :4   Mean   :4   Mean   : 2.524   Mean   :0.359
##  3rd Qu.:123.00   3rd Qu.:6   3rd Qu.:6   3rd Qu.: 3.750   3rd Qu.:0.000
##  Max.   :160.00   Max.   :7   Max.   :7   Max.   :21.000   Max.   :9.000
##                                            NA's   :1
##       prel           negevent          posevent         gender
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :1.000
##  1st Qu.:0.4167   1st Qu.:0.1583   1st Qu.:0.600   1st Qu.:1.000
##  Median :2.0000   Median :0.3500   Median :0.950   Median :2.000
##  Mean   :2.5830   Mean   :0.4414   Mean   :1.048   Mean   :1.562
##  3rd Qu.:4.0000   3rd Qu.:0.6292   3rd Qu.:1.378   3rd Qu.:2.000
##  Max.   :9.0000   Max.   :2.3767   Max.   :3.883   Max.   :2.000
##
##       rosn            age           desired          state
##  Min.   :2.100   Min.   :24.43   Min.   :1.000   Min.   :2.333
##  1st Qu.:3.200   1st Qu.:30.53   1st Qu.:3.333   1st Qu.:3.667
##  Median :3.500   Median :34.57   Median :4.667   Median :4.000
##  Mean   :3.436   Mean   :34.29   Mean   :4.465   Mean   :3.966
##  3rd Qu.:3.800   3rd Qu.:38.19   3rd Qu.:5.667   3rd Qu.:4.222
```

```
##  Max.   :4.000   Max.   :42.28   Max.   :8.000   Max.   :5.000
##                                   NA's   :3       NA's   :3
```

```r
output_variable <- "numall"
input_variables <- setdiff(names(drinks.selected), output_variable)

plots <- lapply(input_variables, function(variable) {
  ggplot(drinks.selected, aes_string(x = variable, y = output_variable)) +
    geom_point() +
    labs(x = variable, y = output_variable) +
    ggtitle(paste(variable, "vs", output_variable))
})

gridExtra::grid.arrange(grobs = plots, ncol = 4)
```



When visually inspecting each of the covariances and scatterplots it appears that there may be a relationship between some of the variables and the number of drinks but it appears as most of them will not be strong indicators of the final number of drinks consumed. `desired` has the strongest relationship with `numall` with a covariance of `0.40`. For some of the variables the pattern seems to defy expectation such as with `rosn` where those with the strongest sense of self-confidence appeared to drink more.

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.

```r
deviance_plot <- function(model) {
  s.res <- rstandard(model, type = "pearson")
  lin.pred <- model$linear.predictors
  df <- data.frame(s.res, lin.pred)
  df %>%
    ggplot(aes(x = df$lin.pred , y = df$s.res)) +
    geom_point() +
    geom_hline(yintercept=c(3, 2, 0, -2, -3), color = "red", linetype = "dashed") +
    geom_smooth(se = FALSE) +
    ggtitle("Standardized residuals") +
    xlab("Linear predictor") +
    ylab("Standardized Pearson residuals")
}

# Poisson
model_nrel_saturday <- glm(formula = numall ~ nrel,
  family = poisson,
  data = saturday
)

summary(model_nrel_saturday)
```

```
##
## Call:
## glm(formula = numall ~ nrel, family = poisson, data = saturday)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.39003    0.05715  24.320   <2e-16 ***
## nrel         0.04971    0.05076   0.979    0.328
## ---
## 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: 249.43  on 87  degrees of freedom
## AIC: 508.83
##
## Number of Fisher Scoring iterations: 5
```

```r
# Quasi-Poisson
model_quasi_nrel_saturday <- glm(formula = numall ~ nrel,
  family = quasipoisson(link = "log"),
  data = saturday
)

summary(model_quasi_nrel_saturday)
```

```
##
## Call:
## glm(formula = numall ~ nrel, family = quasipoisson(link = "log"),
##     data = saturday)
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.39003    0.10114  13.743   <2e-16 ***
## nrel         0.04971    0.08983   0.553    0.581
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 3.131623)
##
##     Null deviance: 250.34  on 88  degrees of freedom
## Residual deviance: 249.43  on 87  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

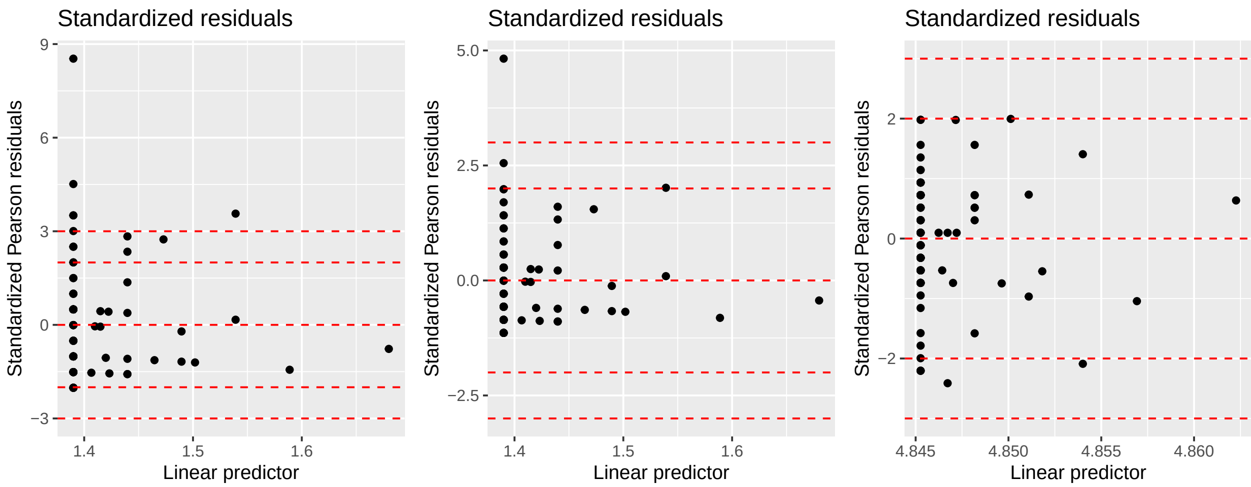
```r
# Desired
model_des_saturday <- glm(formula = desired ~ nrel,
  data = saturday
)

summary(model_des_saturday)
```

```
##
## Call:
## glm(formula = desired ~ nrel, data = saturday)
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.845267   0.184642  26.241   <2e-16 ***
## nrel        0.002914   0.178607   0.016    0.987
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 2.572294)
##
```

```
##     Null deviance: 223.79  on 88  degrees of freedom
## Residual deviance: 223.79  on 87  degrees of freedom
## AIC: 340.64
##
## Number of Fisher Scoring iterations: 2
```

```
deviance_plot(model_nrel_saturday) +
  deviance_plot(model_quasi_nrel_saturday) +
  deviance_plot(model_des_saturday)
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



> By itself, it does `nrel` does not have a significant effect on the number of drinks one has. There does appear to be some possible issues with overdispersion according to the residual plot and the high deviance, so a second model was fitted using a quasi-poisson model. This one did not appear to have issues with overdispersion but also did not indicate a significant relationship between `nrel` and `numall`. A model for `desired` also did not indicate any significant relationship between `nrel` and the desire one has for a drink. Altogether, it does not appear as though one can confirm the hypothesis that `nrel` has an effect on `desired` or `numall`.

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.

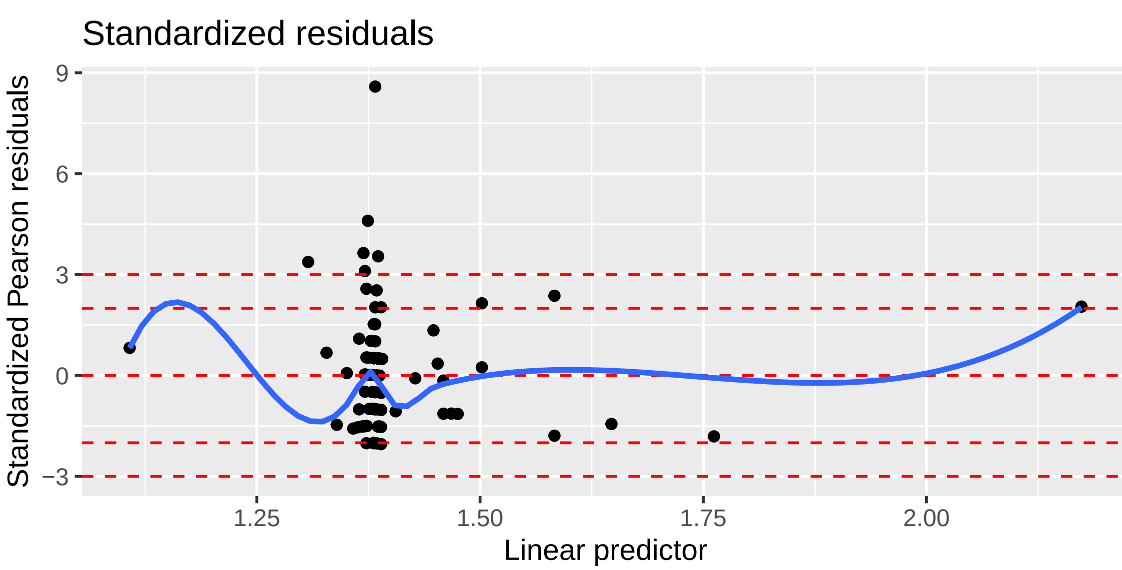
```
model_nrel_rosn_saturday <- glm(formula = numall ~ nrel*rosn,
  family = poisson,
  data = saturday
)

summary(model_nrel_rosn_saturday)
```

```
##
```

```
## Call:
## glm(formula = numall ~ nrel * rosn, family = poisson, data = saturday)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.32343    0.46367   2.854  0.00431 **
## nrel         1.07253    0.45716   2.346  0.01897 *
## rosn         0.01642    0.13403   0.123  0.90248
## nrel:rosn   -0.28731    0.13036  -2.204  0.02752 *
## ---
## 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: 244.30  on 85  degrees of freedom
## AIC: 507.7
##
## Number of Fisher Scoring iterations: 5
```

```
deviance_plot(model_nrel_rosn_saturday)
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



As the researchers predicted it appears that both `nrel` and `nrel:rosn` have a significant effect on `numall`. While the expected number of drinks increases with `nrel`, as `rosn` increases, the effect of `nrel` is dampened. Although, what is considered "high self-esteem" is never specified, it is possible under this model to have a high enough self-esteem that negative relationship interactions no longer have an evident effect on the number of drinks consumed. This confirms the hypothesis the researchers were trying to prove. Additionally, it appears as though adding `rosn` to the variables explains the data well enough that the model does not seem at risk of overdispersion as the previous one was.