W271 Spring 18: Lab 2

Alyssa Eisenberg, Jeffrey Hsu, Gerard Kelly

Alcohol Consumption, Self-Esteem and Romantic Interactions

Section copied from EDA with basic setup

```
library(car); require(dplyr); library(Hmisc); library(mcprofile); library(ggplot2); library(gridExtra);
dehart <- read.table(file="DeHartSimplified.csv", header=TRUE, sep=",")
dehart$dayweek_f <- factor(dehart$dayweek); levels(dehart$dayweek_f) = c("mon","tue","wed","thu","fri",
dehart$gender_f <- factor(dehart$gender); levels(dehart$gender_f) = c("male","female")

dehart_nrel = dehart[which(dehart$nrel != 0),]

dehart$rosn_cat <- cut(dehart$rosn, breaks=c(-1, 2.8, 3.4, Inf), labels = c("low", "mid", "high"))

dehart$trel <- dehart$nrel/sd(dehart$nrel) + dehart$prel/sd(dehart$prel)
```

Additional Data Prep For Modeling

```
# Transform number of negative romantic events into a categorical dummy variable (0 vs any events) beca
# 1 means there have been negative relationship events, 0 means there have not
dehart$nrel_dummy <- as.numeric(dehart$nrel != 0)


# Decide on bins for desire to drink in order to use ordinal logistic regression model
# Generate potential bin values based on distribution - small bin option for each 1 unit interval, larg
dehart$desired_binS <- cut(dehart$desired, breaks=c(-1, 2, 3, 4, 5, 6, 7, 8))
summary(dehart$desired_binS)
```

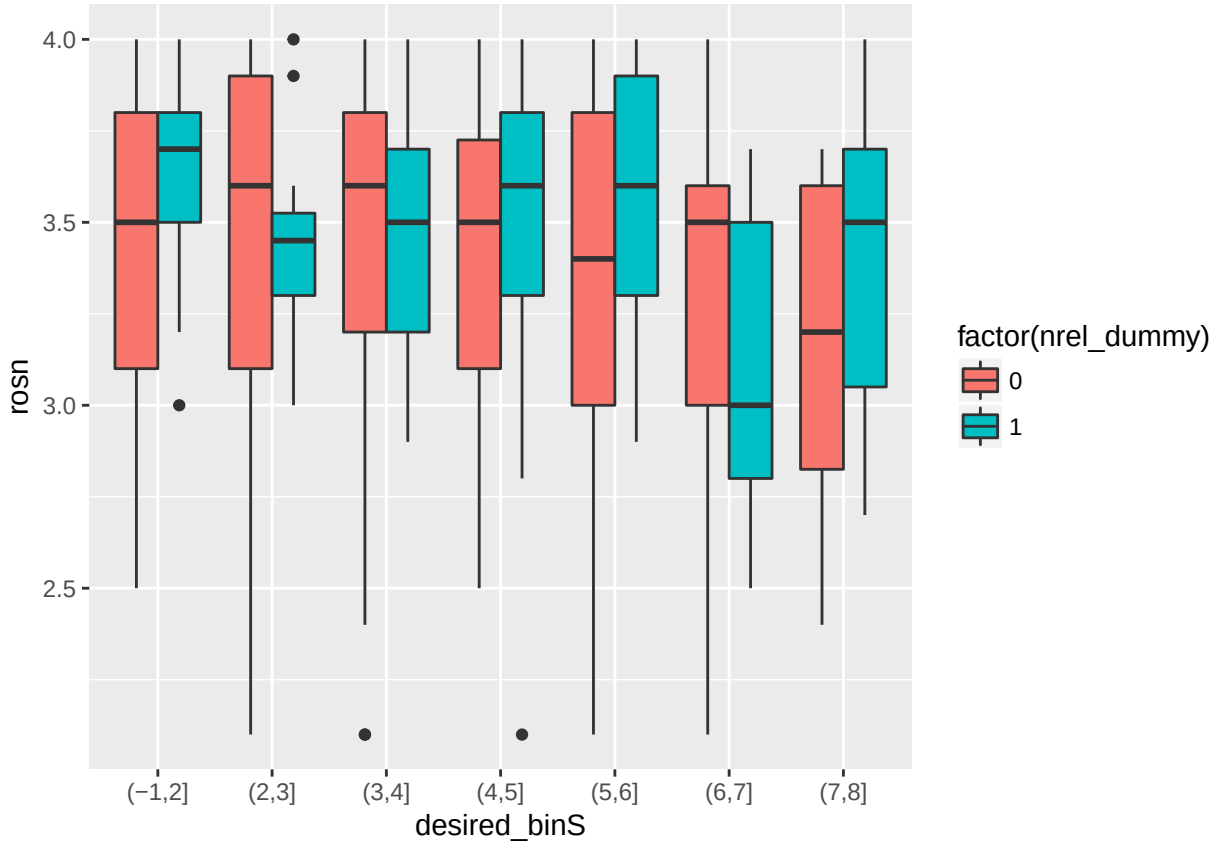
```
## (-1,2]  (2,3]  (3,4]  (4,5]  (5,6]  (6,7]  (7,8]   NA's
##     79     61    104    157    133     54     32      3
```

```
dehart$desired_binL <- cut(dehart$desired, breaks=c(-1, 3.5, 6, 8))
summary(dehart$desired_binL)
```

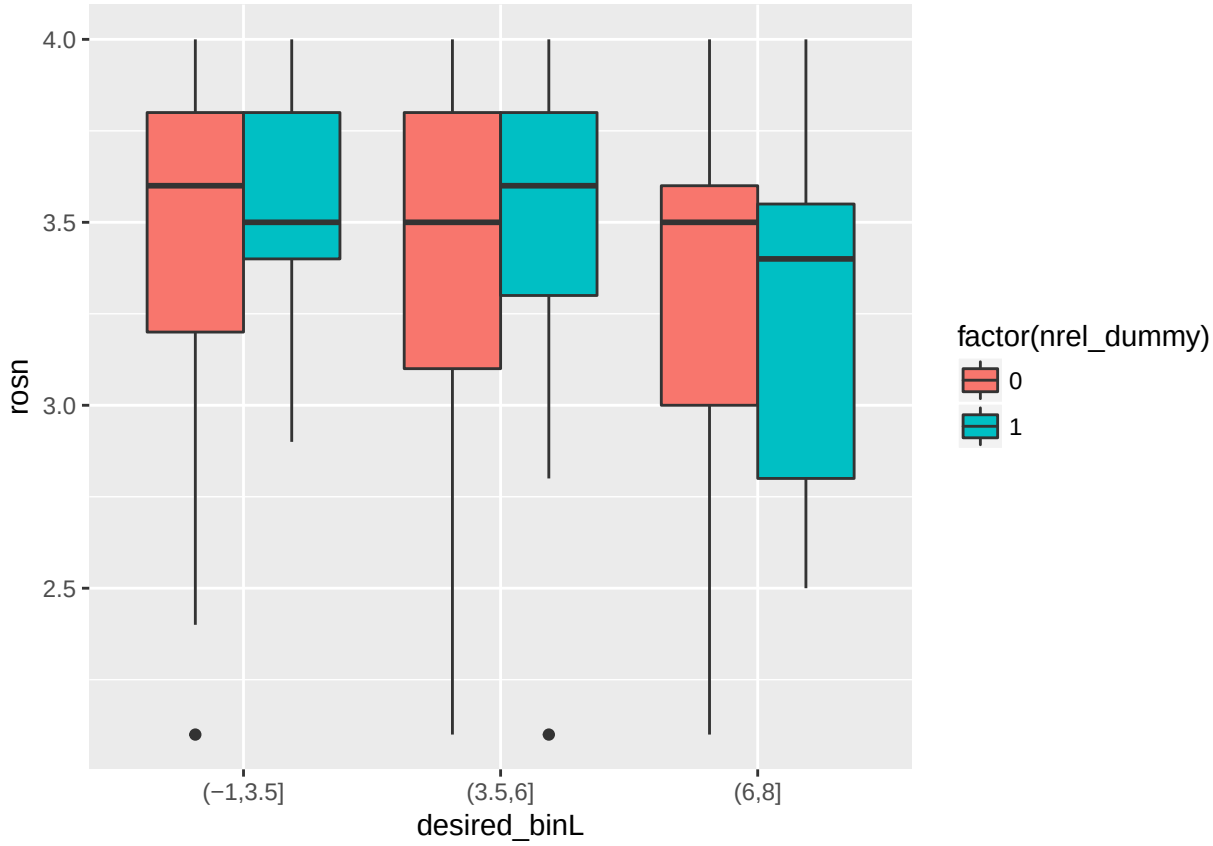
```
## (-1,3.5]  (3.5,6]    (6,8]     NA's
##      162      372       86        3
```

```
# Check presence of negative events and low self esteem in each bin. If not enough variation, may need

ggplot(na.omit(dehart), aes(y=rosn, x=desired_binS, fill=factor(nrel_dummy))) + geom_boxplot()
```




```
ggplot(na.omit(dehart), aes(y=rosn, x=desired_binL, fill=factor(nrel_dummy))) + geom_boxplot()
```




```
# in general seeing that those who have negative romantic relationships tend to not have the very low l
# smaller bins appear too small. Seeing a lot of variation in distributions of self-esteem across those
# We do see that among those with negative interactions, low self-esteem is only showing up with a high

# Remove data points where there is an observation missing (note: do not remove the entire individual)
dehart_clean <- dehart[!(is.na(dehart$numall) | is.na(dehart$state) | is.na(dehart$desired)),]
```

Research Goal

The researchers stated the hypothesis as follow: *"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."*

Poisson Model With Number of Drinks Outcome

```
# Version with all data points, which violates independence assumption

# generate base model: nrel dummy, trait self-esteem, interaction(nrel * trait self-esteem)
pois_base <- glm(formula = numall ~ nrel_dummy + rosn + nrel_dummy:rosn, data=dehart_clean, family=pois
summary(pois_base)
```

```
##
## Call:
## glm(formula = numall ~ nrel_dummy + rosn + nrel_dummy:rosn, family = poisson(link = log),
##     data = dehart_clean)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4504  -1.1137  -0.3339   0.5868   7.2275
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      0.74179    0.23394   3.171  0.00152 **
## nrel_dummy       1.06068    0.52477   2.021  0.04326 *
## rosn             0.04959    0.06768   0.733  0.46375
## nrel_dummy:rosn -0.29204    0.15154  -1.927  0.05397 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1583.5  on 617  degrees of freedom
## Residual deviance: 1579.1  on 614  degrees of freedom
## AIC: 2949
##
## Number of Fisher Scoring iterations: 5
```

```
# generate intermediate model: add DOW, prel, interaction(prel * trait self-esteem)
pois <- glm(formula = numall ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn + dayweek_f, data=
summary(pois)
```

```
##
## Call:
## glm(formula = numall ~ nrel_dummy + rosn + nrel_dummy:rosn +
##     prel + prel:rosn + dayweek_f, family = poisson(link = log),
##     data = dehart_clean)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9441  -1.5269  -0.3093   0.5790   6.4810
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -0.38948    0.36720  -1.061 0.288835
## nrel_dummy       1.50872    0.54022   2.793 0.005226 **
## rosn             0.28495    0.10439   2.730 0.006341 **
## prel             0.26661    0.07520   3.545 0.000392 ***
## dayweek_ftue    -0.13910    0.11068  -1.257 0.208836
## dayweek_fwed    -0.07122    0.10783  -0.661 0.508923
## dayweek_fthu     0.20405    0.10136   2.013 0.044109 *
## dayweek_ffri     0.37870    0.09755   3.882 0.000104 ***
## dayweek_fsat     0.67950    0.09222   7.368 1.73e-13 ***
## dayweek_fsun     0.19079    0.10200   1.870 0.061430 .
## nrel_dummy:rosn -0.40743    0.15578  -2.615 0.008910 **
## rosn:prel       -0.06765    0.02224  -3.041 0.002354 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1583.5  on 617  degrees of freedom
## Residual deviance: 1425.0  on 606  degrees of freedom
## AIC: 2810.9
##
## Number of Fisher Scoring iterations: 5
```

```r
# generate full model: add age, gender, negevent, posevent, state self-esteem
pois_full <- glm(formula = numall ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn + dayweek_f +
summary(pois_full)
```

```
##
## Call:
## glm(formula = numall ~ nrel_dummy + rosn + nrel_dummy:rosn +
##     prel + prel:rosn + dayweek_f + age + gender_f + negevent +
##     posevent + state, family = poisson(link = log), data = dehart_clean)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.1541  -1.5030  -0.3386   0.5897   6.8990
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -0.079431   0.467098  -0.170  0.86497
## nrel_dummy       1.320795   0.548939   2.406  0.01612 *
## rosn             0.328918   0.108503   3.031  0.00243 **
## prel             0.259333   0.078401   3.308  0.00094 ***
## dayweek_ftue    -0.125458   0.110788  -1.132  0.25746
## dayweek_fwed    -0.042230   0.108211  -0.390  0.69635
## dayweek_fthu     0.219894   0.101466   2.167  0.03022 *
## dayweek_ffri     0.386206   0.097640   3.955 7.64e-05 ***
## dayweek_fsat     0.687339   0.092385   7.440 1.01e-13 ***
## dayweek_fsun     0.186452   0.102306   1.822  0.06838 .
## age              0.001748   0.005812   0.301  0.76354
## gender_ffemale  -0.124407   0.053098  -2.343  0.01913 *
## negevent        -0.213315   0.076945  -2.772  0.00557 **
## posevent         0.069871   0.046258   1.510  0.13092
## state           -0.112374   0.061779  -1.819  0.06892 .
## nrel_dummy:rosn -0.334642   0.159168  -2.102  0.03551 *
## rosn:prel       -0.067254   0.022943  -2.931  0.00337 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1583.5  on 617  degrees of freedom
## Residual deviance: 1407.8  on 601  degrees of freedom
## AIC: 2803.8
##
## Number of Fisher Scoring iterations: 5
```

```r
# LRT for the models
anova(pois_base, pois, test="Chisq") #p-value practically 0, reject null that they explain the same amo
```

```
## Analysis of Deviance Table
##
## Model 1: numall ~ nrel_dummy + rosn + nrel_dummy:rosn
## Model 2: numall ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn +
##     dayweek_f
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1       614     1579.1
## 2       606     1425.0  8   154.11 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(pois, pois_full, test="Chisq") #p-value=0.4%, reject null that they explain the same amount of va
```

```
## Analysis of Deviance Table
##
## Model 1: numall ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn +
##     dayweek_f
## Model 2: numall ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn +
##     dayweek_f + age + gender_f + negevent + posevent + state
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       606     1425.0
## 2       601     1407.8  5   17.174 0.004182 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
Anova(pois, test="LR") # indicates that DOW and prel variables are significant
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: numall
##                 LR Chisq Df Pr(>Chisq)
## nrel_dummy         2.559  1  0.1096809
## rosn               0.054  1  0.8156706
## prel              13.777  1  0.0002059 ***
## dayweek_f        117.624  6  < 2.2e-16 ***
## nrel_dummy:rosn    6.706  1  0.0096106 **
## rosn:prel          9.145  1  0.0024942 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
Anova(pois_full, test="LR") # indicates that gender and negevent are significant
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: numall
##                 LR Chisq Df Pr(>Chisq)
## nrel_dummy         6.668  1   0.009818 **
## rosn               0.476  1   0.490092
## prel               6.166  1   0.013022 *
## dayweek_f        115.088  6  < 2.2e-16 ***
## age                0.091  1   0.763475
## gender_f           5.480  1   0.019232 *
## negevent           7.902  1   0.004938 **
```

```
## posevent           2.248  1   0.133762
## state              3.277  1   0.070245 .
## nrel_dummy:rosn    4.350  1   0.037018 *
## rosn:prel          8.470  1   0.003610 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# specifically test prel and interaction
pois_test3 <- glm(formula = numall ~ nrel_dummy + rosn + nrel_dummy:rosn + dayweek_f, data=dehart_clean
anova(pois_test3, pois, test="Chisq") #p-value practically 0, prel and interaction are significant
```

```
## Analysis of Deviance Table
##
## Model 1: numall ~ nrel_dummy + rosn + nrel_dummy:rosn + dayweek_f
## Model 2: numall ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn +
##     dayweek_f
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1       608     1447.9
## 2       606     1425.0  2   22.922 1.053e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# test model including gender and negevent, and then test negevent interaction
pois_test1 <- glm(formula = numall ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn + dayweek_f
pois_test2 <- glm(formula = numall ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn + dayweek_f

anova(pois, pois_test1, test="Chisq") # p-value 0.3%, reject null that they explain same variance
```

```
## Analysis of Deviance Table
##
## Model 1: numall ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn +
##     dayweek_f
## Model 2: numall ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn +
##     dayweek_f + gender_f + negevent
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       606       1425
## 2       604       1413  2   11.961 0.002527 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(pois_test1, pois_test2, test="Chisq") # p-value 59%, fail to reject null that they explain same v
```

```
## Analysis of Deviance Table
##
## Model 1: numall ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn +
##     dayweek_f + gender_f + negevent
## Model 2: numall ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn +
##     dayweek_f + gender_f + negevent + negevent:rosn
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       604     1413.0
## 2       603     1412.7  1  0.29584   0.5865
```

```
# SUGGESTED FINAL MODEL
pois_final <- glm(formula = numall ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn + dayweek_f
summary(pois_final)
```

```
##
```

```
## Call:
## glm(formula = numall ~ nrel_dummy + rosn + nrel_dummy:rosn +
##     prel + prel:rosn + dayweek_f + gender_f + negevent, family = poisson(link = log),
##     data = dehart_clean)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.0984  -1.4836  -0.3180   0.5749   6.7944
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -0.28621    0.36851  -0.777 0.437352
## nrel_dummy       1.31264    0.54570   2.405 0.016154 *
## rosn             0.28760    0.10459   2.750 0.005963 **
## prel             0.25469    0.07575   3.362 0.000773 ***
## dayweek_ftue    -0.13409    0.11069  -1.211 0.225774
## dayweek_fwed    -0.05675    0.10800  -0.525 0.599261
## dayweek_fthu     0.21100    0.10135   2.082 0.037349 *
## dayweek_ffri     0.37873    0.09754   3.883 0.000103 ***
## dayweek_fsat     0.67644    0.09217   7.339 2.15e-13 ***
## dayweek_fsun     0.17895    0.10210   1.753 0.079668 .
## gender_ffemale  -0.11011    0.05239  -2.102 0.035565 *
## negevent        -0.18181    0.07485  -2.429 0.015147 *
## nrel_dummy:rosn -0.32948    0.15814  -2.083 0.037210 *
## rosn:prel       -0.06343    0.02242  -2.829 0.004671 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1583.5  on 617  degrees of freedom
## Residual deviance: 1413.0  on 604  degrees of freedom
## AIC: 2803
##
## Number of Fisher Scoring iterations: 5
```

```
#for final model, test nrel and interaction
pois_final_testInt <- glm(formula = numall ~ nrel_dummy + rosn + prel + prel:rosn + dayweek_f + gender_
pois_final_testNrel <- glm(formula = numall ~ rosn + prel + prel:rosn + dayweek_f + gender_f + negevent
anova(pois_final_testInt, pois_final, test="Chisq") #p-value = 3.9%, interaction term is signficant
```

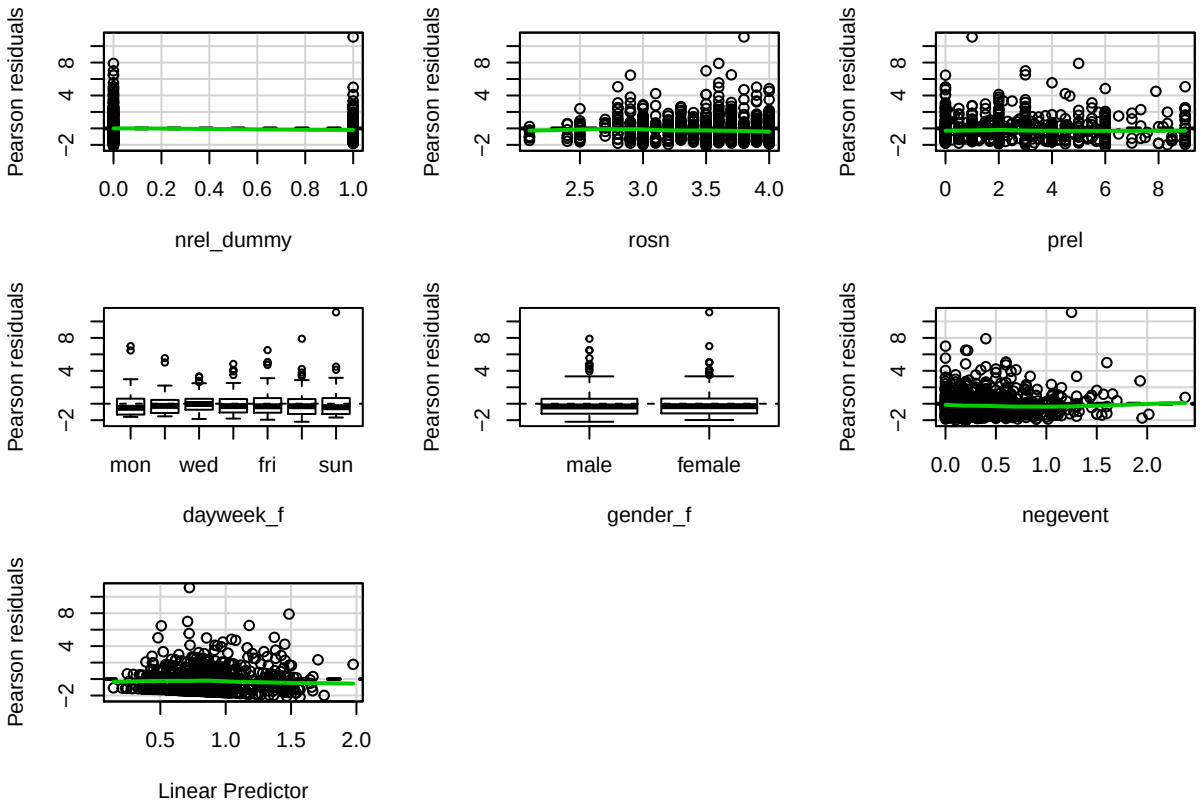
```
## Analysis of Deviance Table
##
## Model 1: numall ~ nrel_dummy + rosn + prel + prel:rosn + dayweek_f + gender_f +
##     negevent
## Model 2: numall ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn +
##     dayweek_f + gender_f + negevent
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       605     1417.3
## 2       604     1413.0  1   4.2693  0.03881 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(pois_final_testNrel, pois_final, test="Chisq") #p-value=0.3%, nrel and interaction jointly signif
```

```
## Analysis of Deviance Table
##
## Model 1: numall ~ rosn + prel + prel:rosn + dayweek_f + gender_f + negevent
## Model 2: numall ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn +
##     dayweek_f + gender_f + negevent
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       606     1424.8
## 2       604     1413.0  2   11.726 0.002843 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# residual plots for final model

suppressWarnings(residualPlots(pois_final, layout=c(3,3)))
```




```
##             Test stat Pr(>|t|)
## nrel_dummy      0.000    1.000
## rosn            2.155    0.142
## prel            0.498    0.481
## dayweek_f          NA       NA
## gender_f           NA       NA
## negevent        2.387    0.122
```

```r
# In Pearson residual plots vs. explanatory variables, looking for: same variance throughout range of r
# In Pearson residual plots vs. fitted values, looking for: same variance and no fluctuation in mean. T
```

```
# In Pearson residual plots vs. linear predictor, looking for: same variance and no fluctuation in mean
# All of the above plots, looking for: extreme residuals. Only about 5% should be beyond abs(2), typica
# Test stats only make sense for numeric variables. Null hypothesis is that there is a relationship btw


# Run final model once for each day and compare against version with all data
```

Run ordinal logistic regression model with Desire to Drink Outcome

Version with all data points, which violates independence assumption

```
library(MASS)
```

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

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

```
#Version with all data points, which violates independence assumption

# generate base model: nrel dummy, trait self-esteem, interaction(nrel * trait self-esteem)
prop_odds_base <- polr(formula = desired_binL ~ nrel_dummy + rosn + nrel_dummy:rosn, data=dehart_clean,
summary(prop_odds_base)
```

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

```
## Call:
## polr(formula = desired_binL ~ nrel_dummy + rosn + nrel_dummy:rosn,
##     data = dehart_clean, method = "logistic")
##
## Coefficients:
##                    Value Std. Error t value
## nrel_dummy        2.5266     1.7414   1.451
## rosn             -0.6145     0.2141  -2.870
## nrel_dummy:rosn -0.6636     0.4975  -1.334
##
## Intercepts:
##                   Value   Std. Error t value
## (-1,3.5]|(3.5,6] -3.1250  0.7478    -4.1788
## (3.5,6]|(6,8]    -0.2014  0.7336    -0.2746
##
## Residual Deviance: 1135.801
## AIC: 1145.801
```

```
# generate intermediate model: add DOW, prel, interaction(prel * trait self-esteem)
prop_odds <- polr(formula = desired_binL ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn + day
summary(prop_odds)
```

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

```
## Call:
## polr(formula = desired_binL ~ nrel_dummy + rosn + nrel_dummy:rosn +
```

```
##     prel + prel:rosn + dayweek_f, data = dehart_clean, method = "logistic")
##
## Coefficients:
##                     Value Std. Error t value
## nrel_dummy        2.97239    1.79192  1.6588
## rosn             -0.40919    0.30989 -1.3204
## prel              0.32608    0.25415  1.2830
## dayweek_ftue      0.50131    0.30124  1.6641
## dayweek_fwed      0.57487    0.30289  1.8979
## dayweek_fthu      0.56540    0.30405  1.8595
## dayweek_ffri      0.88706    0.30369  2.9209
## dayweek_fsat      0.90870    0.30474  2.9819
## dayweek_fsun     -0.11438    0.30226 -0.3784
## nrel_dummy:rosn -0.76941    0.51131 -1.5048
## rosn:prel       -0.07505    0.07436 -1.0093
##
## Intercepts:
##                   Value   Std. Error t value
## (-1,3.5]|(3.5,6] -1.7821  1.0823    -1.6466
## (3.5,6]|(6,8]     1.2473  1.0804     1.1545
##
## Residual Deviance: 1109.06
## AIC: 1135.06
```

```
# generate full model: add age, gender, negevent, posevent, state self-esteem
prop_odds_full <- polr(formula = desired_binL ~ nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn +
summary(prop_odds_full)
```

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

```
## Call:
## polr(formula = desired_binL ~ nrel_dummy + rosn + nrel_dummy:rosn +
##     prel + prel:rosn + dayweek_f + age + gender_f + negevent +
##     posevent + state, data = dehart_clean, method = "logistic")
##
## Coefficients:
##                     Value Std. Error t value
## nrel_dummy        2.79016    1.81544  1.5369
## rosn             -0.50849    0.32573 -1.5611
## prel              0.25392    0.26414  0.9613
## dayweek_ftue      0.49179    0.30299  1.6231
## dayweek_fwed      0.55845    0.30518  1.8299
## dayweek_fthu      0.58149    0.30554  1.9031
## dayweek_ffri      0.87652    0.30478  2.8759
## dayweek_fsat      0.88783    0.30625  2.8991
## dayweek_fsun     -0.15367    0.30464 -0.5044
## age              -0.00362    0.01829 -0.1979
## gender_ffemale   -0.39569    0.17157 -2.3062
## negevent          0.06330    0.23705  0.2670
## posevent          0.20175    0.15555  1.2970
## state             0.35323    0.20093  1.7580
## nrel_dummy:rosn -0.69486    0.51974 -1.3369
## rosn:prel       -0.06473    0.07672 -0.8438
##
```

```
## Intercepts:
##                   Value   Std. Error t value
## (-1,3.5]|(3.5,6] -0.9271  1.4283    -0.6491
## (3.5,6]|(6,8]     2.1404  1.4308     1.4959
##
## Residual Deviance: 1099.151
## AIC: 1135.151
```

```
# LRT for the models
anova(prop_odds_base, prop_odds, test="Chisq") #p-value=.07%, reject null that they explain the same am
```

```
## Likelihood ratio tests of ordinal regression models
##
## Response: desired_binL
##                                                                 Model
## 1                                nrel_dummy + rosn + nrel_dummy:rosn
## 2 nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn + dayweek_f
##   Resid. df Resid. Dev   Test    Df LR stat.      Pr(Chi)
## 1       613   1135.801
## 2       605   1109.060 1 vs 2     8 26.74115 0.0007835306
```

```
anova(prop_odds, prop_odds_full, test="Chisq") #p-value=7.8%, fail to reject at 5% level that they expl
```

```
## Likelihood ratio tests of ordinal regression models
##
## Response: desired_binL
##
## 1                                                nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:r
## 2 nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn + dayweek_f + age + gender_f + negevent + p
##   Resid. df Resid. Dev   Test    Df LR stat.    Pr(Chi)
## 1       605   1109.060
## 2       600   1099.151 1 vs 2     5 9.909274 0.07784711
```

```
# Looking at whether prel is needed in intermediate model.
# generate test model: only DOW, no prel
prop_odds_test1 <- polr(formula = desired_binL ~ nrel_dummy + rosn + nrel_dummy:rosn + dayweek_f, data=
prop_odds_test2 <- polr(formula = desired_binL ~ nrel_dummy + rosn + nrel_dummy:rosn + dayweek_f + prel

anova(prop_odds_test1, prop_odds, test="Chisq") #p-value=6.9%, fails to reject at 5% level that prel an
```

```
## Likelihood ratio tests of ordinal regression models
##
## Response: desired_binL
##                                                                 Model
## 1                    nrel_dummy + rosn + nrel_dummy:rosn + dayweek_f
## 2 nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn + dayweek_f
##   Resid. df Resid. Dev   Test    Df LR stat.    Pr(Chi)
## 1       607   1114.403
## 2       605   1109.060 1 vs 2     2 5.343409 0.06913428
```

```
anova(prop_odds_test2, prop_odds, test="Chisq") # p-value=31%, fail to reject that prel interaction exp
```

```
## Likelihood ratio tests of ordinal regression models
##
## Response: desired_binL
##                                                                 Model
## 1             nrel_dummy + rosn + nrel_dummy:rosn + dayweek_f + prel
```

```
## 2 nrel_dummy + rosn + nrel_dummy:rosn + prel + prel:rosn + dayweek_f
##   Resid. df Resid. Dev   Test    Df LR stat.   Pr(Chi)
## 1       606   1110.079
## 2       605   1109.060 1 vs 2     1 1.018839 0.3127946
```

```
anova(prop_odds_test1, prop_odds_test2, test="Chisq") # p-value=3.8%, reject null that they explain sam
```

```
## Likelihood ratio tests of ordinal regression models
##
## Response: desired_binL
##                                                      Model Resid. df
## 1        nrel_dummy + rosn + nrel_dummy:rosn + dayweek_f       607
## 2 nrel_dummy + rosn + nrel_dummy:rosn + dayweek_f + prel       606
##   Resid. Dev   Test    Df LR stat.   Pr(Chi)
## 1   1114.403
## 2   1110.079 1 vs 2     1  4.32457 0.0375659
```

```
Anova(prop_odds, test="LR") #DOW fixed effects significant, and prel significant at 5% level
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: desired_binL
##                 LR Chisq Df Pr(>Chisq)
## nrel_dummy        2.1869  1  0.1391875
## rosn             15.0031  1  0.0001073 ***
## prel              4.3246  1  0.0375659 *
## dayweek_f        20.4766  6  0.0022770 **
## nrel_dummy:rosn   2.2503  1  0.1335922
## rosn:prel         1.0188  1  0.3127946
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Indicates to me that our model should include prel but not interaction



# Looking at whether any of the full model covariates should be included in final model.
Anova(prop_odds_full, test="LR") # gender appears significant
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: desired_binL
##                 LR Chisq Df Pr(>Chisq)
## nrel_dummy        3.1716  1   0.074928 .
## rosn             15.3158  1  9.095e-05 ***
## prel              0.6449  1   0.421931
## dayweek_f        20.7059  6   0.002072 **
## age               0.0392  1   0.843145
## gender_f          5.3549  1   0.020664 *
## negevent          0.0713  1   0.789449
## posevent          1.6931  1   0.193198
## state             3.0943  1   0.078568 .
## nrel_dummy:rosn   1.7785  1   0.182336
## rosn:prel         0.7124  1   0.398644
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
prop_odds_test3 <- polr(formula = desired_binL ~ nrel_dummy + rosn + nrel_dummy:rosn + dayweek_f + prel
anova(prop_odds_test2, prop_odds_test3, test="Chisq") #p-value=3.4%, significant at 5% level to reject
```

```
## Likelihood ratio tests of ordinal regression models
##
## Response: desired_binL
##                                                                 Model
## 1            nrel_dummy + rosn + nrel_dummy:rosn + dayweek_f + prel
## 2 nrel_dummy + rosn + nrel_dummy:rosn + dayweek_f + prel + gender_f
##   Resid. df Resid. Dev   Test    Df LR stat.    Pr(Chi)
## 1       606   1110.079
## 2       605   1105.580 1 vs 2     1  4.49891 0.03391646
```

```r
# Indicates to me that we should include gender


# SUGGESTED FINAL MODEL
prop_odds_final <- polr(formula = desired_binL ~ nrel_dummy + rosn + nrel_dummy:rosn + dayweek_f + prel
summary(prop_odds_final)
```

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

```
## Call:
## polr(formula = desired_binL ~ nrel_dummy + rosn + nrel_dummy:rosn +
##     dayweek_f + prel + gender_f, data = dehart_clean, method = "logistic")
##
## Coefficients:
##                     Value Std. Error t value
## nrel_dummy        2.61700    1.78685  1.4646
## rosn             -0.58764    0.21665 -2.7124
## dayweek_ftue      0.51143    0.30204  1.6933
## dayweek_fwed      0.57421    0.30343  1.8924
## dayweek_fthu      0.57785    0.30459  1.8972
## dayweek_ffri      0.89458    0.30422  2.9406
## dayweek_fsat      0.91745    0.30541  3.0040
## dayweek_fsun     -0.10698    0.30316 -0.3529
## prel              0.07173    0.03478  2.0624
## gender_ffemale   -0.35171    0.16630 -2.1149
## nrel_dummy:rosn -0.64983    0.50986 -1.2745
##
## Intercepts:
##                   Value   Std. Error t value
## (-1,3.5]|(3.5,6] -2.5621  0.7821    -3.2761
## (3.5,6]|(6,8]     0.4807  0.7727     0.6221
##
## Residual Deviance: 1105.58
## AIC: 1131.58
```

```r
#for final model, test nrel and interaction
prop_odds_final_testInt <- polr(formula = desired_binL ~ nrel_dummy + rosn + dayweek_f + prel + gender_
prop_odds_final_testNrel <- polr(formula = desired_binL ~ rosn + dayweek_f + prel + gender_f, data=dehar
anova(prop_odds_final_testInt, prop_odds_final, test="Chisq") #p-value = 20%, interaction term is not s
```

```
## Likelihood ratio tests of ordinal regression models
##
```

```
## Response: desired_binL
##                                                                 Model
## 1                   nrel_dummy + rosn + dayweek_f + prel + gender_f
## 2 nrel_dummy + rosn + nrel_dummy:rosn + dayweek_f + prel + gender_f
##   Resid. df Resid. Dev   Test    Df LR stat.   Pr(Chi)
## 1       606   1107.192
## 2       605   1105.580 1 vs 2     1 1.611898 0.2042253
```

```
anova(prop_odds_final_testNrel, prop_odds_final, test="Chisq") #p-value=9%, nrel and interaction not jo
```

```
## Likelihood ratio tests of ordinal regression models
##
## Response: desired_binL
##                                                                 Model
## 1                                rosn + dayweek_f + prel + gender_f
## 2 nrel_dummy + rosn + nrel_dummy:rosn + dayweek_f + prel + gender_f
##   Resid. df Resid. Dev   Test    Df LR stat.    Pr(Chi)
## 1       607   1110.358
## 2       605   1105.580 1 vs 2     2 4.777726 0.09173392
```

```
# residual plots for final model??




# Run final model once for each day and compare against version with all data
```

```
# Comparison of initial and final models

library(stargazer)
```

```
##
## Please cite as:
##  Hlavac, Marek (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2. http://CRAN.R-project.org/package=stargazer
```

```
stargazer(pois_base,pois, pois_final, pois_full, prop_odds_base, prop_odds, prop_odds_final, prop_odds_
          star.cutoffs = c(.05,.01,.001),
          header=F
          #,type="text"
          )
```

** REMAINING QUESTIONS: - are there other model diagnostics we need to do? - Do we need robust standard errors? - Need to add explanation of coefficients of interest. Do we also want CI or graphs of some sort?

Table 1:

	Dependent variable:							
	numall					desired_binL		
	Poisson					*ordered logistic*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
nrel_dummy	1.061*	1.509**	1.313*	1.321*	2.527	2.972	2.617	2.79
	(0.525)	(0.540)	(0.546)	(0.549)	(1.741)	(1.792)	(1.787)	(1.81
rosn	0.050	0.285**	0.288**	0.329**	−0.614**	−0.409	−0.588**	−0.5
	(0.068)	(0.104)	(0.105)	(0.109)	(0.214)	(0.310)	(0.217)	(0.32
prel		0.267***	0.255***	0.259***		0.326	0.072*	0.25
		(0.075)	(0.076)	(0.078)		(0.254)	(0.035)	(0.26
dayweek_ftue		−0.139	−0.134	−0.125		0.501	0.511	0.49
		(0.111)	(0.111)	(0.111)		(0.301)	(0.302)	(0.30
dayweek_fwed		−0.071	−0.057	−0.042		0.575	0.574	0.55
		(0.108)	(0.108)	(0.108)		(0.303)	(0.303)	(0.30
dayweek_fthu		0.204*	0.211*	0.220*		0.565	0.578	0.58
		(0.101)	(0.101)	(0.101)		(0.304)	(0.305)	(0.30
dayweek_ffri		0.379***	0.379***	0.386***		0.887**	0.895**	0.877
		(0.098)	(0.098)	(0.098)		(0.304)	(0.304)	(0.30
dayweek_fsat		0.679***	0.676***	0.687***		0.909**	0.917**	0.888
		(0.092)	(0.092)	(0.092)		(0.305)	(0.305)	(0.30
dayweek_fsun		0.191	0.179	0.186		−0.114	−0.107	−0.1
		(0.102)	(0.102)	(0.102)		(0.302)	(0.303)	(0.30
age				0.002				−0.0
				(0.006)				(0.01
gender_ffemale			−0.110*	−0.124*			−0.352*	−0.39
			(0.052)	(0.053)			(0.166)	(0.17
negevent			−0.182*	−0.213**				0.06
			(0.075)	(0.077)				(0.23
posevent				0.070				0.20
				(0.046)				(0.15
state				−0.112				0.35
				(0.062)				(0.20
nrel_dummy:rosn	−0.292	−0.407**	−0.329*	−0.335*	−0.664	−0.769	−0.650	−0.6
	(0.152)	(0.156)	(0.158)	(0.159)	(0.498)	(0.511)	(0.510)	(0.52
rosn:prel		−0.068**	−0.063**	−0.067**		−0.075		−0.0
		(0.022)	(0.022)	(0.023)		(0.074)		(0.07
Constant	0.742**	−0.389	−0.286	−0.079				
	(0.234)	(0.367)	(0.369)	(0.467)				