Count Distribution Project

Chris Corona

4/14/23

Introduction

Heavy drinking and negative self-thoughts have long been conjectured to be associated behaviors. People with low self-esteem and perceived negative interactions in their lives are believed to be more susceptible to heavy drinking habits and vice versa. In the study by DeHart 2008, "moderate to heavy drinkers" (at least 12 alcoholic drinks/week for women, 15 for men) were recruited to keep a daily record of each drink that they consumed. Participants also completed a variety of rating scales covering daily events in their lives and items related to self-esteem. The research question for this report hypothesizes that negative interactions with romantic partners would be associated with alcohol consumption. We predict 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.

Data

This data set follows 89 different participants for 7 consecutive days, making 623 total observations. However 5 individuals had missing values for a few important variables, so their entire set of observations were discarded. After removal of these observations, we have 84 participants for 7 consecutive days, making 588 total observations. 13 different variables were measured, 4 categorical and 9 quantitative. The variable of interest is the number of drinks consumed. There are also several demographic variables and measures of self-esteem. All variable descriptions are below as well as summary statistics in Table 1.

`id`: participant id

`studyday`: day in the study, all participants are tracked for 7 days

`dayweek`: day of the week

`drinks`: number of drinks consumed on the given day in the study

`nrel`: index for combining the total number and intensity of negative relationship events experienced during the day (higher values indicate more negative events)

`prel`: index for combining the total number and intensity of positive relationship events experienced during the day (higher values indicate more positive events)

`negevent`: index for combining the total number and intensity of negative non-relationship events experienced during the day (higher values indicate more negative events)

`posevent`: index for combining the total number and intensity of positive non-relationship events experienced during the day (higher values indicate more positive events)

`sex`: "Male" or "Female"

`rosn`: Rosenberg (1965) 10-item self-esteem scale for global self-evaluation; higher scores indicate higher "trait self-esteem"

`age`: age (years)

`desired`: daily desire to drink; higher scores indicate greater desire to drink

`state`: Heatherton & Polivy (1991) 9-item scale for participants' "state self-esteem"; higher scores indicate higher "state self-esteem"

```
##
##                       age   desired   drinks   negevent     nrel   posevent
## --------------- -------- --------- -------- ---------- -------- ----------
##            Mean    34.17      4.47     2.54       0.44     0.35       1.04
##         Std.Dev     4.53      1.71     2.70       0.38     0.93       0.64
##             Min    24.43      1.00     0.00       0.00     0.00       0.00
##          Median    34.38      4.67     2.00       0.35     0.00       0.94
##             Max    42.28      8.00    21.00       2.02     9.00       3.88
##         N.Valid   588.00    588.00   588.00     588.00   588.00     588.00
##       Pct.Valid   100.00    100.00   100.00     100.00   100.00     100.00
##
## Table: Table continues below
##
##
##
##                      prel     rosn    state
## --------------- -------- -------- --------
##            Mean     2.56     3.42     3.95
##         Std.Dev     2.39     0.42     0.44
##             Min     0.00     2.10     2.33
##          Median     2.00     3.50     4.00
##             Max     9.00     4.00     5.00
##         N.Valid   588.00   588.00   588.00
##       Pct.Valid   100.00   100.00   100.00
```

Table 1: Summary statistics for the variables in the data set

We begin with an exploratory data analysis to understand the structure and relationships between variables. Figure 1 clearly shows that the response variable, number of drinks consumed, comes from a count distribution. Since the number of drinks consumed does not have an upper bound, we can assume that the Poisson or negative binomial distributions might be a good representation of the count data. However the count distribution changes for different days of the week, shown in Figure 2. The most notable difference is between Monday and Saturday. We will explore this difference later. Figure 3 shows that none of the quantitative variables have much correlation with the response, except the daily desire to drink. The research question wants to determine the nature of the relationship between number of drinks consumed and the index of negative relationship events. Figure 4 plots these two variables with a linear regression fit on top. It is unclear from this plot if there is any relationship there. We will explore this more in the next section. The research question also hypothesizes about an interaction between the index of negative relationship events and trait self-esteem. Here we decided to treat the trait self-esteem variable as categorical, creating five bins of equal length splitting the range of values evenly. Figure 5 shows panels for the five bins of trait self-esteem, plotting drinks vs index of negative relationship events. This series of plots suggests there might be an interaction present since there appears to be different slopes for the five different bins of trait self-esteem, with the most positive slope corresponding to the lowest trait self-esteem and the most negative slope corresponding to the highest trait self-esteem. This interaction will also be explored in more detail in the next section. Figure 6 shows panels for the first nine participants in the data set, plotting their distribution of drinks. Each individual can have a drastically different distribution.

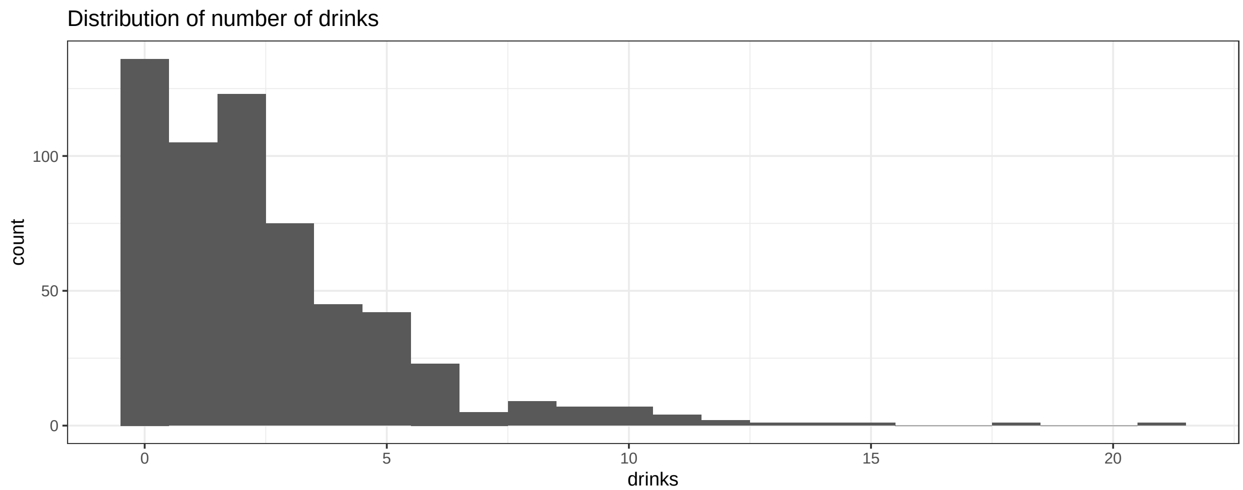

Figure 1: Distribution of number of drinks

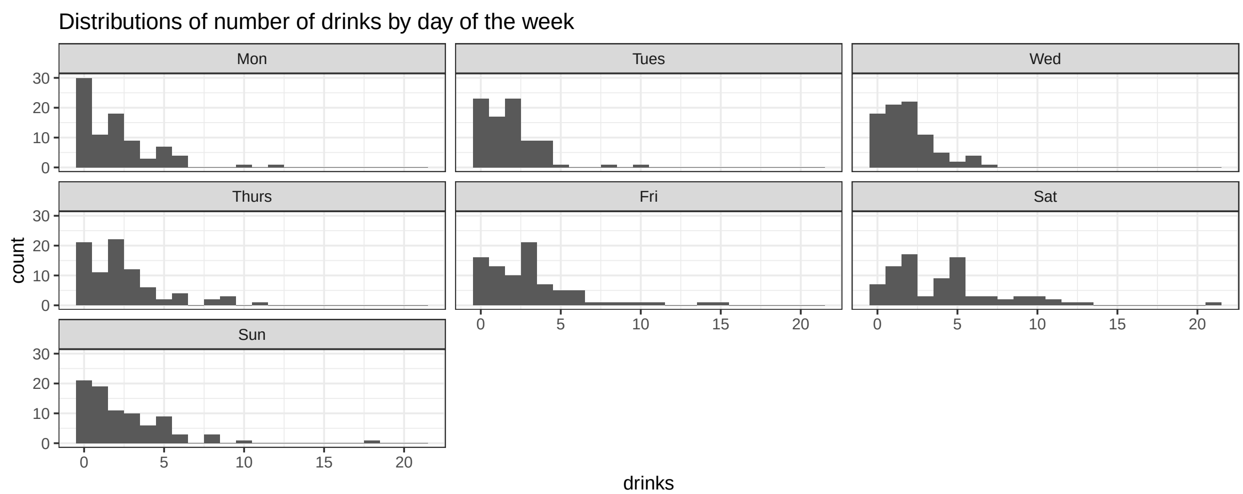

Figure 2: Distributions of number of drinks by day of the week

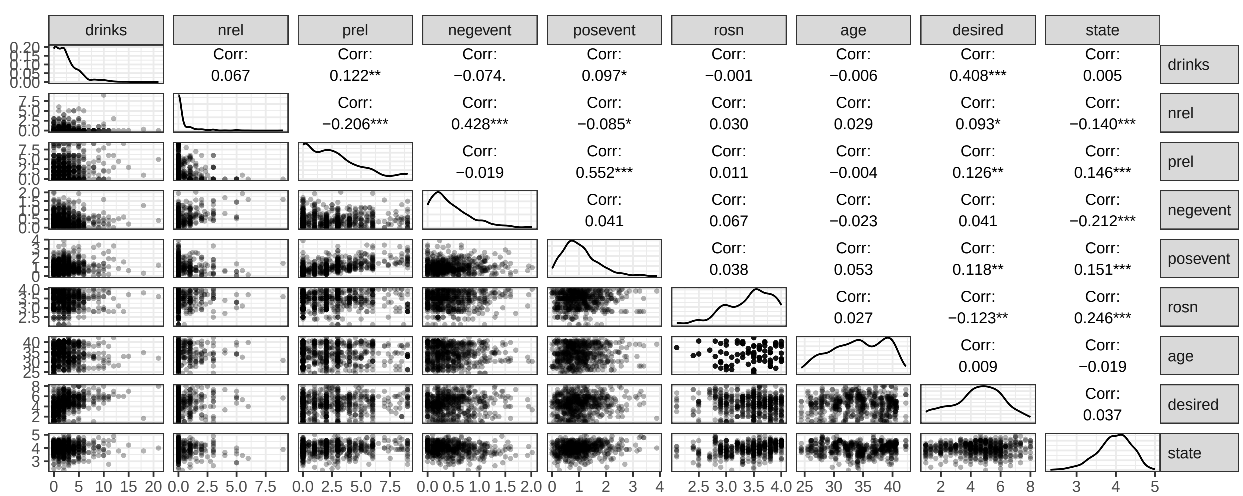

Figure 3: Correlation matrix between all quantitative variables in the data set

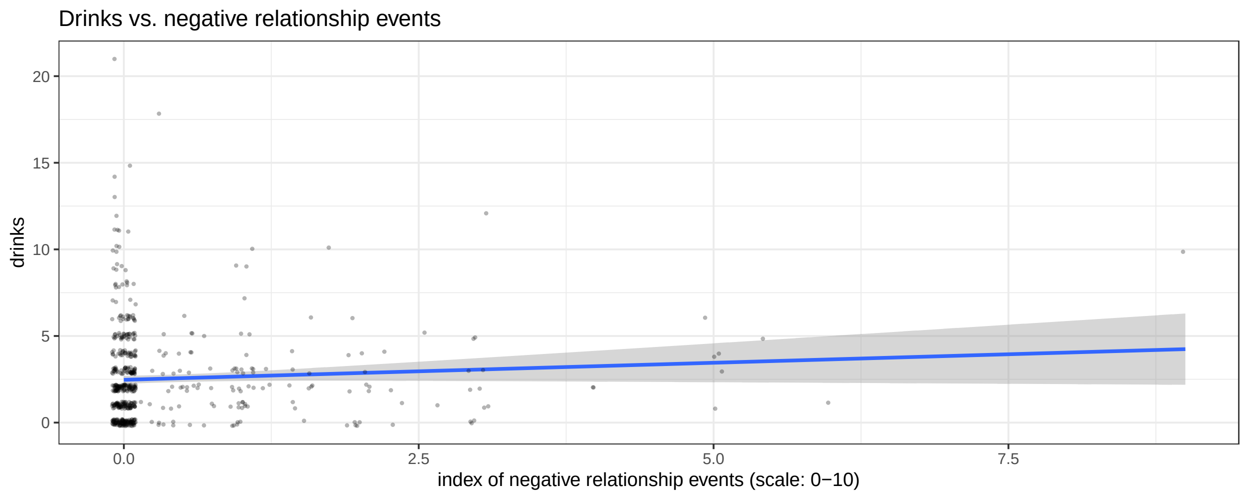

Figure 4: Plot of drinks vs index of negative relationship events with linear regression fit

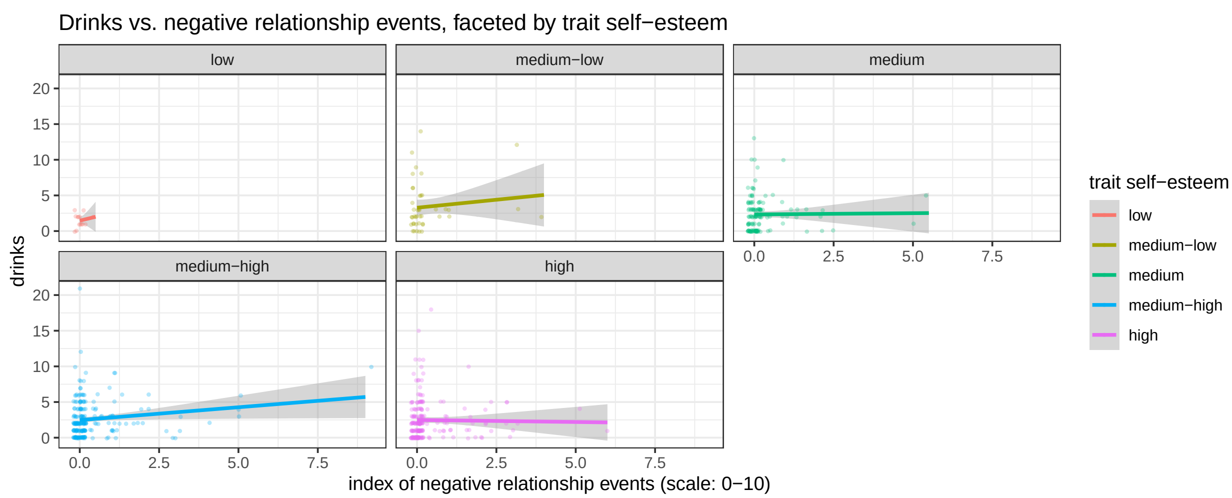

Figure 5: Plot of drinks vs index of negative relationship events with linear regression fit, faceted by trait self-esteem

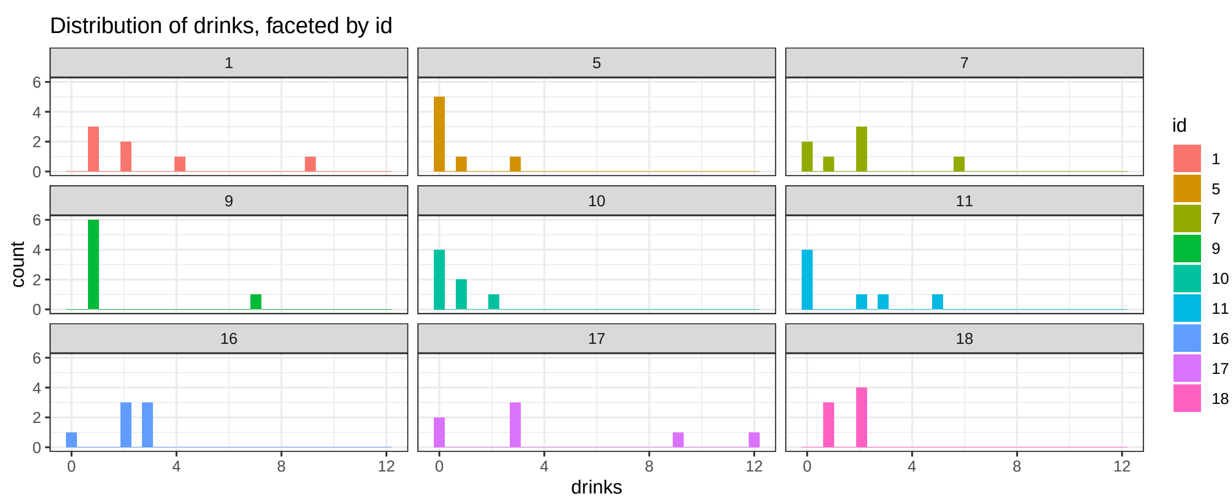

Figure 6: Distribution of drinks, faceted by id for the first 9 participants

Now we will find the most appropriate count distribution to model this data. Let's consider the data collected on Mondays, shown in Figure 7. We construct probability mass functions (PMFs) from Poisson and negative binomial distributions and compare them to this distribution of the number of drinks for Monday. To estimate the parameters for each of the distributions, we fit a generalized linear model of the corresponding type to the Monday data. Because there are repeat observations by each participant in the study for each day of the week, the models we fit account for the individual as a random effect. For the Poisson distribution, we need to estimate λ which is simply the mean estimated response. For the negative binomial, we need to estimate both the mean and dispersion parameters. The mean comes directly from the glm estimated response. The dispersion is estimated with $\hat{r} = \frac{\bar{x}^2}{\bar{s}^2 - \bar{x}^2}$. Both the estimated Poisson and negative binomial distributions are plotted in Figure 6. From this plot, the estimated Poisson distribution appears to be a better fit than the estimated negative binomial. The Poisson estimate is closer to the data tha the negative binomial estimate for all counts. Therefore we will model the number of drinks with a Poisson distribution. Table 2 shows the parameter estimates for the Poisson and negative binomial distributions for number of drinks on Monday.

model	mean	dispersion
Poisson	0.307	
Negative Binomial	0.307	0.019

Table 2: Estimated parameters for the Poisson and negative binomial distributions

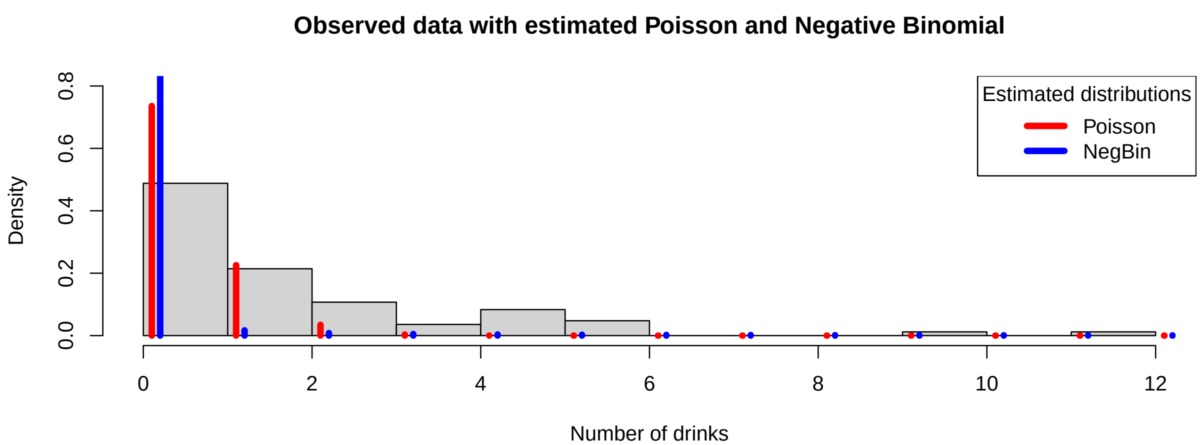

Figure 7: Distribution of the number of drinks for Monday and the estimated Poisson and negative binomial distributions for this data

We will also look at whether the distribution of the number of drinks depends on day of the week. As mentioned above, the two most different distributions are those of Monday and Saturday as shown in Figure 8, and these are the two that we will be comparing. Since each individual reports a value for both Monday and Saturday, we need to account for this violation of independence between samples. A paired t-test can account for these repeated measures and will determine if the mean difference is significant. In this case, each individual's response for Monday and Saturday will be the pair which will be differenced. The null hypothesis and alternative hypothesis for this paired t-test are the following:

H_0: There is no difference in the distribution between drinks consumed by individuals on Monday vs Saturday for participants in this study

H_A: The is a difference in the distribution between drinks consumed by individuals on Monday vs Saturday for participants in this study

After differencing, we apply a Shapiro-Wilk test to determine if the differences are normal. The differences fail the Shapiro-Wilk test, rejecting that the differences are normal with a p-value of 0.00026. This violates the assumption of normality for the paired t-test. Therefore we will use a paired samples Wilcoxon signed rank test, a non-parametric test, which relaxes this assumption of normality. The Wilcoxon signed rank test results in a p-value of 6e-7 as seen in Table 3. This is strong evidence against the null hypothesis. We reject the null hypothesis that there is no difference in the distribution between drinks consumed by individuals on Monday vs Saturday for participants in this study. The Wilcoxon signed rank test is not comparing the mean difference, but instead the mean signed rank. Therefore we say that the distributions differ in their ordering and thus their medians (not means).

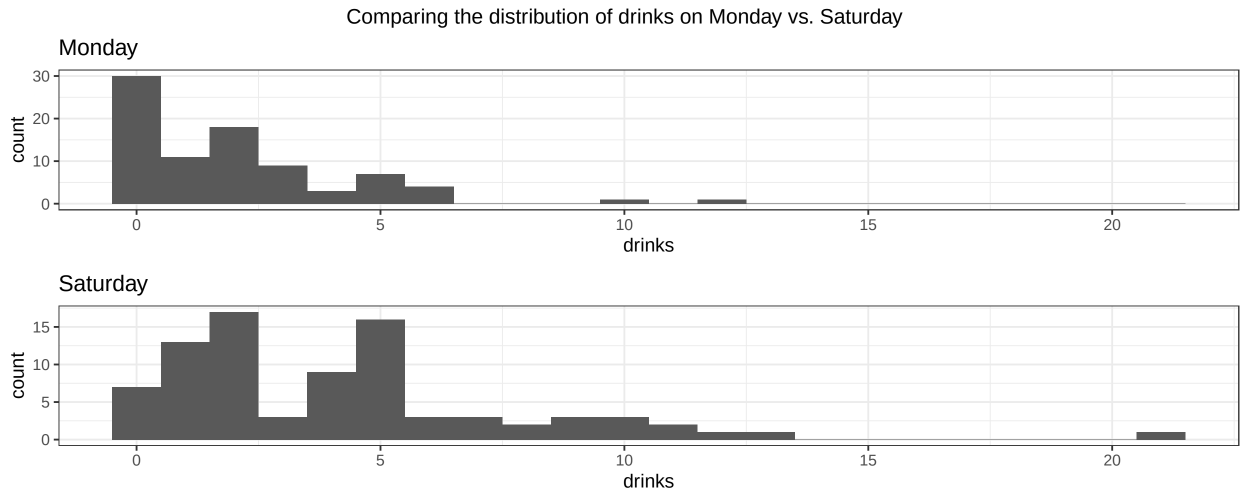

Figure 8: Comparison of the distribution of drinks on Monday vs. Saturday

	V	p.value
V	390.5	6e-07

Table 3: Results from paired samples Wilcoxon signed rank test

Statistical Procedures

In this section, we describe the models used to investigate the research question of whether negative interactions with romantic partners would be associated with alcohol consumption. We predict 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. Previously, we decided that a Poisson model does a better job estimating the data than a negative binomial for drinks consumed on Monday. We replicated that method again, but this time using all days of the week instead of just Monday. We plot the estimated Poisson and negative binomial distributions on top of the observed data in Figure 9. Again the Poisson appears to estimate the data better, so we choose the Poisson model over the negative binomial. To account for repeated observations from individuals, we include the individual as a random effect in the model. Because the research question asks whether there is a relationship between number of drinks consumed and the index of negative relationship events, and whether that effect is different for low vs high trait self-esteem, we will test two models. The first model includes the index of negative relationship events:

$$
\begin{aligned}
&y_i \sim Poisson(\lambda_i),\\
&\eta_i = \beta_0 + \beta_1 X_{nrel} + \alpha_{l[i]},\\
&\text{where } l \text{ is the individual for observation i}\\
&\alpha_l \sim N(0, \sigma^2_\alpha),\\
&\log(\lambda_i) = \eta_i
\end{aligned}
$$

The second model includes the interaction between the index of negative relationship events and trait self-esteem:

$$
\begin{aligned}
&y_i \sim Poisson(\lambda_i),\\
&\eta_i = \beta_0 + \beta_1 X_{nrel} + \sum_{j=2}^{5} \beta_j I(X_{rosn} = level_j) + \sum_{k=6}^{9} \beta_k X_{nrel} \times I(X_{rosn} = level_k) + \alpha_{l[i]},\\
&\text{where } l \text{ is the individual for observation i}\\
&\alpha_l \sim N(0, \sigma^2_\alpha),\\
&\log(\lambda_i) = \eta_i
\end{aligned}
$$

Figure 10 and Figure 11 show the diagnostic plots for the two models. The diagnostics for both models look very similar. There appears to be some slight pattern (maybe non-linear) in the deviance vs fitted plot which indicates there could be some covariate that is missing from our models. The variance appears to be constant, so there is no clear violation of that assumption. And the deviance residuals follow the normal distribution within acceptable bounds.

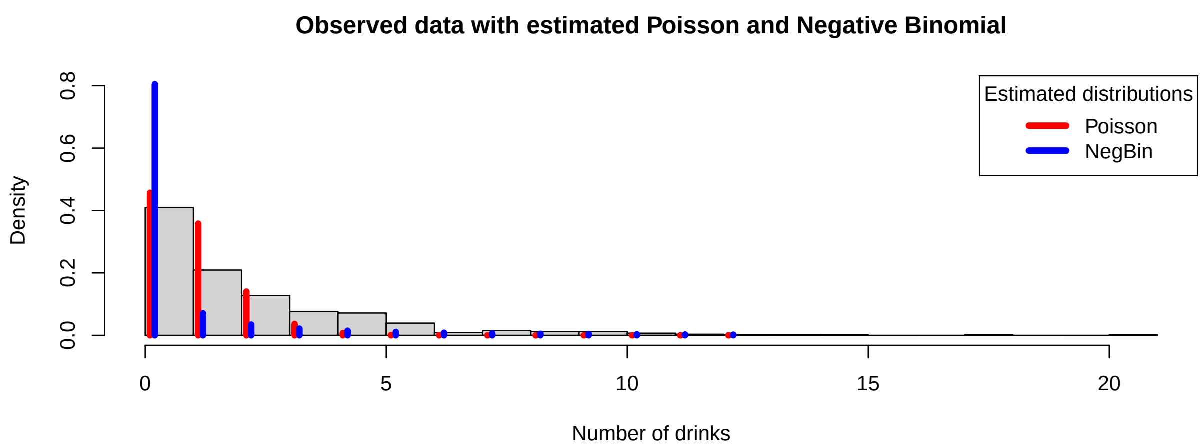

Figure 9: Distribution of the number of drinks for all days of the week and the estimated Poisson and negative binomial distributions for this data

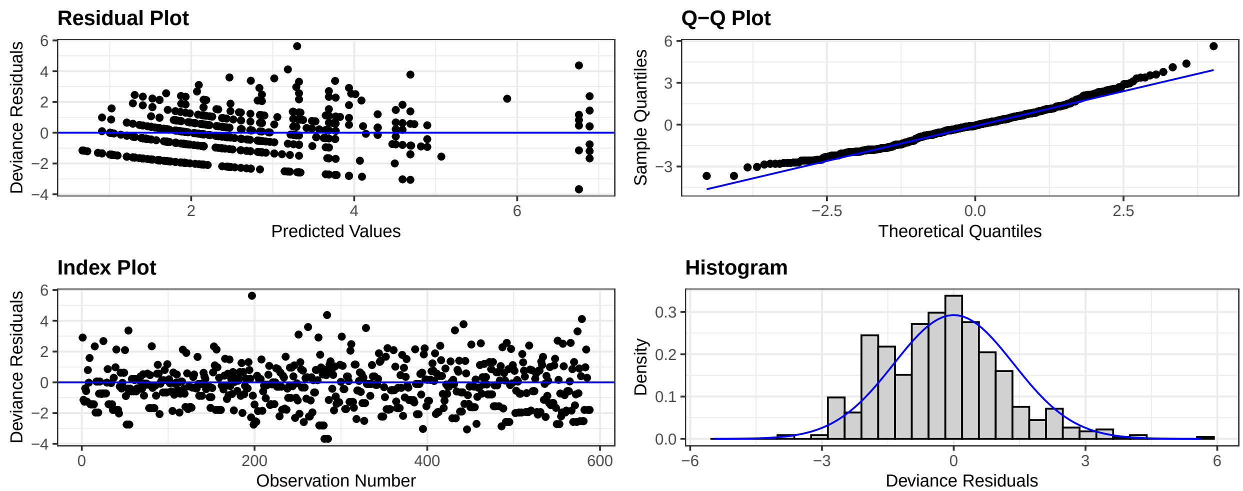

Figure 10: Diagnostic plots for the Poisson mixed effect model including main effect from the index of negative relationship events as fixed effect, individual as random effect

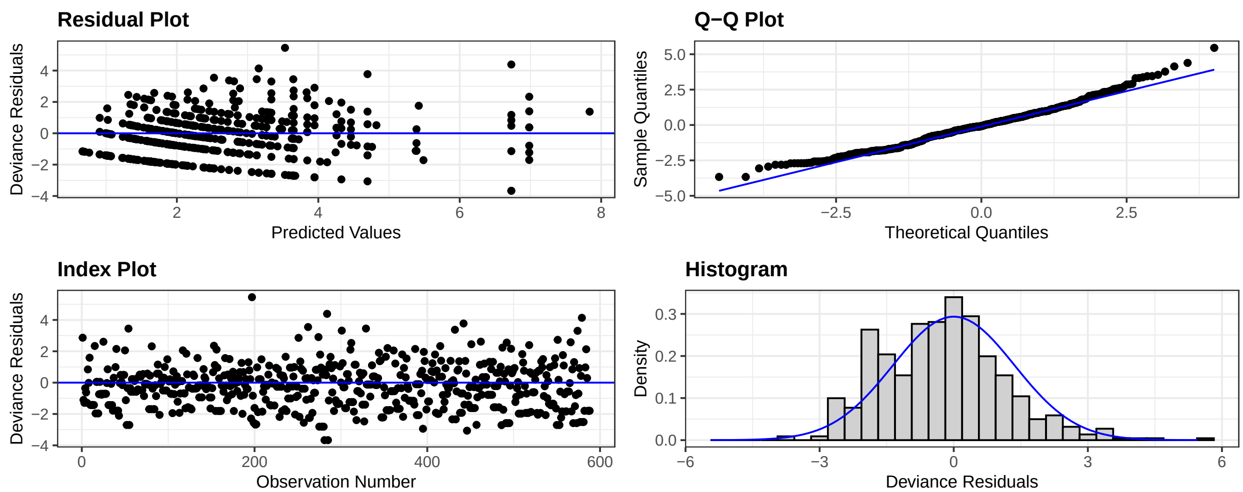

Figure 11: Diagnostic plots for the Poisson mixed effect model including main effects and interaction from the index of negative relationship events and trait self-esteem as fixed effects, individual as random effect

Results

In this section, we predict the point estimates and corresponding 95% confidence interval for each model. In order to obtain these estimates, we first produce the fitted value and associated standard error. To compute the confidence interval for this model, we calculate the lower and upper bounds using the standard normal approximation of $\pm 1.96SE$. Thus, we find the lower and upper bounds correspond to $\hat{y}_i \pm 1.96\hat{SE}_i$.

We fit both mixed effect Poisson models: the model with just the index of negative relationship events as the fixed effect, and the model with the main effects and interaction between the index of negative relationship events and trait self-esteem as the fixed effects. Let's start with the no-interaction model. Table 4 shows that the number of drinks consumed is associated with the index of negative relationship events with a p-value 0.005. The Poisson generalized linear model uses a log link function, so we can apply the inverse to these estimates for interpretation. The days reported as scoring a 0 for the index of negative relationship events, participants in the study consume 2.2 drinks on average with a lower bound of 1.9 and an upper bound of 2.5. For each unit increase in the index of negative relationship events, participants in the study consume another 1.09 drinks on average with a lower bound of 1.03 and an upper bound of 1.15. This first hypothesis that negative relationship events would be associated with alcohol consumption shows strong evidence of rejecting the null hypothesis that there is no association.

Now let's look at the interaction model. Table 5 shows that the number of drinks consumed is not associated with either of the main effects nor their interaction with all p-values greater than 0.05. This second hypothesis that people with low trait self-esteem would drink more on days they experienced more negative relationship events compared with days during which they experienced fewer negative relationship events fails to reject the null hypothesis that there is no interaction.

Figure 12 shows the actual distribution of drinks consumed and the estimated Poisson models. The Poisson model with no interaction has a better fit to the data. Figure 13 shows a plot of the predicted drinks consumed vs index of negative relationship events for the no interaction model. Figure 14 shows a series of plots of the predicted drinks consumed vs index of negative relationship events with a panel for each bin of trait self-esteem for the interaction model. Both Figure 13 and Figure 14 show fits that are similar to the original data, as plotted in the exploratory data analysis above. The scope of inference is for participants in this study since they were not random samples, but instead convenience samples of volunteers.

fixed.effect	estimate	Std.Error	z.value	p.value
intercept	0.783	0.062	12.608	0.000
nrel	0.082	0.029	2.786	0.005

Table 4: Results from Poisson mixed effect model including main effect from the index of negative relationship events as fixed effect, individual as random effect

fixed.effect	estimate	Std.Error	z.value	p.value
intercept	0.349	0.410	0.851	0.395
nrel	0.444	1.518	0.292	0.770
rosn.medium-low	0.645	0.466	1.384	0.166
rosn.medium	0.384	0.430	0.893	0.372
rosn.medium-high	0.429	0.422	1.016	0.310
rosn.high	0.477	0.423	1.127	0.260
nrel:rosn.medium-low	-0.246	1.520	-0.162	0.872
nrel:rosn.medium	-0.388	1.519	-0.255	0.799
nrel:rosn.medium-high	-0.314	1.518	-0.207	0.836
nrel:rosn.high	-0.489	1.519	-0.322	0.747

Table 5: Results from Poisson mixed effect model including main effects and interaction from the index of negative relationship events and trait self-esteem as fixed effects, individual as random effect

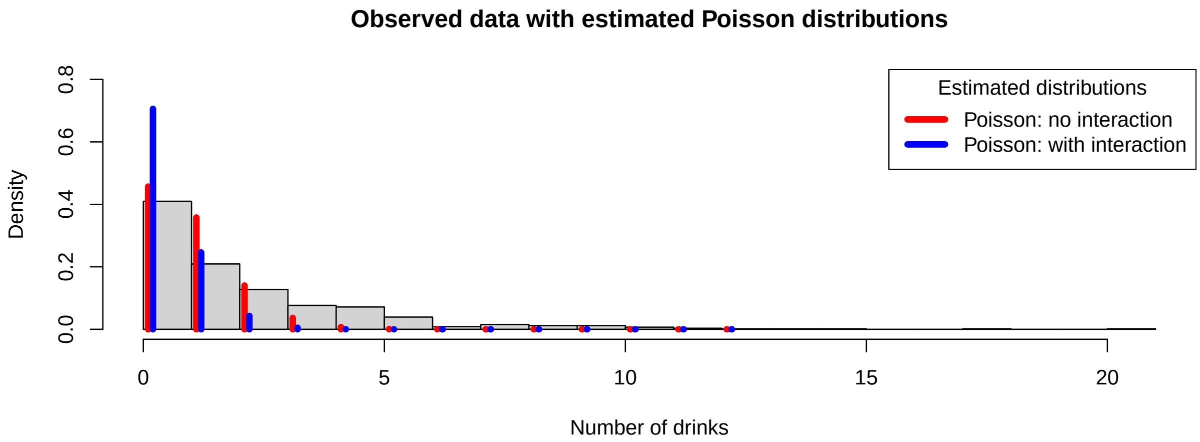

Figure 12: Distribution of the number of drinks for all days of the week and the estimated Poisson distributions for this data

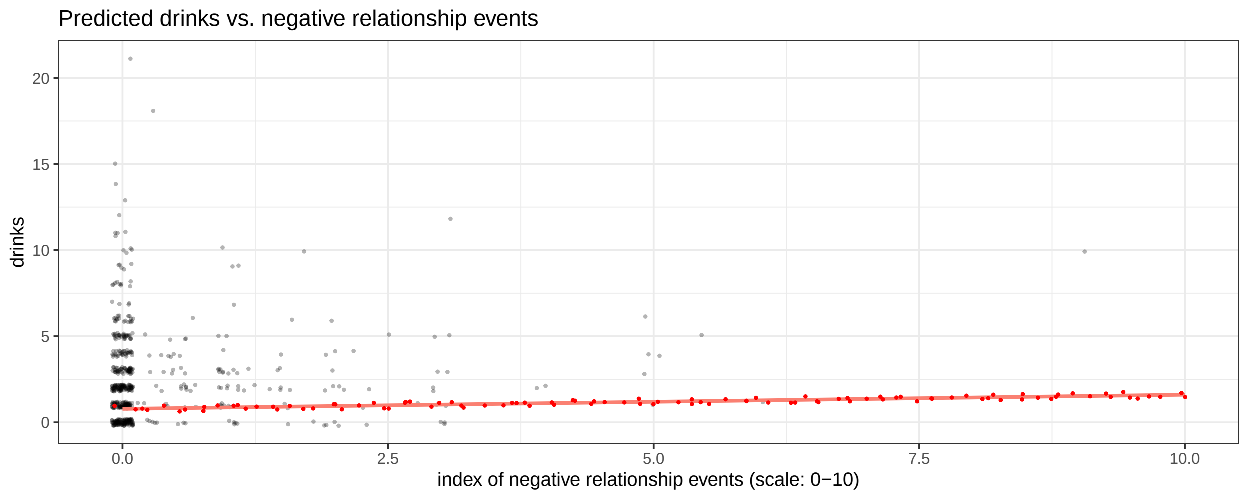

Figure 13: Plot of predicted values for Poisson mixed effect model including main effect from the index of negative relationship events as fixed effect, individual as random effect. Predicted values and regression fit in red, actual values in black.

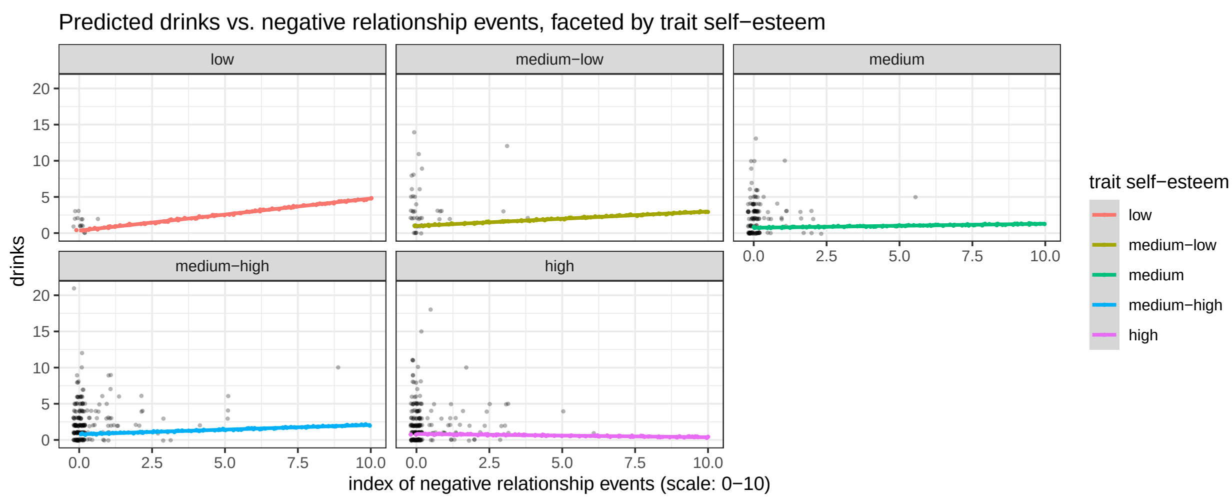

Figure 14: Plot of predicted values for Poisson mixed effect model including main effects and interaction

from the index of negative relationship events and trait self-esteem as fixed effects, individual as random effect. Predicted values and regression fits in color, actual values in black.

Citations

Tracy DeHart, Howard Tennen, Stephen Armeli, Michael Todd, Glenn Affleck, Drinking to regulate negative romantic relationship interactions: The moderating role of self-esteem, Journal of Experimental Social Psychology, Volume 44, Issue 3, 2008, Pages 527-538, ISSN 0022-1031, https://doi.org/10.1016/j.jesp.2007.10.001.

Paired samples Wilcoxon Test in R. STHDA. (n.d.). Retrieved April 14, 2023, from http://www.sthda.com/english/wiki/paired-samples-wilcoxon-test-in-r

Appendix

```
# read in data
dat <- read_csv("DeHartSimplified.csv")

dat$id <- factor(dat$id)
dat$gender <- factor(dat$gender, levels = c(1, 2), labels = c("Male", "Female"))
dat$dayweek <- factor(dat$dayweek, levels = 1:7, labels = c('Mon','Tues','Wed','Thurs','Fri', 'Sat','Sun
dat <- dat %>% rename(sex = gender, drinks = numall)
na_ids <- dat %>% filter(is.na(state) | is.na(desired) | is.na(drinks)) %>% dplyr::select(id) %>% pull()

dehart <- dat %>% filter(!id %in% na_ids)
dehart$id <- factor(dehart$id)

library(knitr)
library(summarytools)
descr(dehart[-c(1:2)],
      headings = FALSE,
      stats = "common",
      split.tables=80)

ggplot(data=dehart, aes(x=drinks)) +
  geom_histogram(binwidth=1) +
  ggtitle("Distribution of number of drinks")

ggplot(data=dehart, aes(x=drinks)) +
  geom_histogram(binwidth=1) +
  facet_wrap(~dayweek) +
  ggtitle("Distributions of number of drinks by day of the week")

library(GGally)
ggpairs(dehart[-c(1:3,9)],
        upper = list(continuous = wrap(ggally_cor,
                                       size = 3,
                                       color ="black")),
        lower = list(continuous = wrap("points", alpha = 0.3,    size=0.75))) +
  theme(strip.text.y = element_text(angle=0, hjust=0))

# plot drinks vs nrel main effect
```

```r
ggplot(data=dehart, aes(x=nrel, y=drinks)) +
  geom_smooth(method="lm") +
  geom_jitter(width=0.1, height=0.2, alpha=0.3, size=0.5) +
  labs(title="Drinks vs. negative relationship events",
       y="drinks",
       x="index of negative relationship events (scale: 0-10)")

# preliminary variables for auto-binning
n_bins = 5
range = max(dehart$rosn) - min(dehart$rosn)
# bins by fixed values (same width of bins)
dehart <- dehart %>% mutate(trait_self_esteem = cut(dehart$rosn,
                                       breaks = seq(from=min(dehart$rosn),
                                                    to=max(dehart$rosn),
                                                    by=range/n_bins),
                                       include.lowest=TRUE,
                                       labels=c("low","medium-low","medium","medium-high","high")))

# plot drinks vs nrel with rosn interaction effects
ggplot(data=dehart, aes(x=nrel, y=drinks, color=trait_self_esteem)) +
  geom_smooth(method="lm") +
  facet_wrap(~ trait_self_esteem) +
  geom_jitter(width=0.2, height=0.1, alpha=0.3, size=0.5) +
  labs(title="Drinks vs. negative relationship events, faceted by trait self-esteem",
       y="drinks",
       x="index of negative relationship events (scale: 0-10)",
       color= "trait self-esteem") +
  scale_color_discrete(labels=c("low","medium-low","medium","medium-high","high"))

# plot drinks faceted by id
ggplot(data=dehart[c(1:63),], aes(x=drinks, fill=id)) +
  geom_histogram() +
  facet_wrap(~ id) +
  labs(title="Distribution of drinks, faceted by id",
       y="count",
       x="drinks",
       fill= "id")

library(MASS)
reduced_dehart <- dehart %>% filter(dayweek %in% c("Mon",'Sat'))
mon <- reduced_dehart %>% filter(dayweek == 'Mon')

# fit poisson mixed effect model, accounting for person as random effect
m.pois <- glmer(drinks ~ (1 | id), family="poisson", data=mon)
sum.pois <- summary(m.pois)

# fit negative binomial mixed effect model, accounting for person as random effect
m.nb <- glmer.nb(drinks ~ (1 | id), data=mon)
sum.nb <- summary(m.nb)

# estimate parameters
lambda.pois <- sum.pois$coefficients[1]
mean.nb <- sum.nb$coefficients[1]
```

```r
dispersion.nb <- mean.nb^2/(var(mon$drinks)-mean.nb)

df.params <- data.frame(model=c("Poisson","Negative Binomial"),
                        mean=c(round(lambda.pois,3),round(mean.nb,3)),
                        dispersion=c("",round(dispersion.nb,3)))
kable(df.params)

# generate poisson/nb from estimated parameters
x <- c(0:12)
y_pois <- dpois(x, lambda=lambda.pois)
y_nb <- dnbinom(x, size=dispersion.nb, mu=mean.nb)

# plot results
hist(mon$drinks, freq=F, breaks=12, main="Observed data with estimated Poisson and Negative Binomial", 
points(x+0.1, y_pois, col="red", type="h", lwd=5)
points(x+0.2, y_nb, col="blue", type="h", lwd=5)
legend("topright", title="Estimated distributions", legend=c("Poisson", "NegBin"), col=c("red", "blue")

library(mosaic)
library(stats)
library(gridExtra)

p1 <- ggplot(data=reduced_dehart[reduced_dehart$dayweek=="Mon",], aes(x=drinks)) +
  geom_histogram(binwidth=1) +
  xlim(-0.5,21.5) +
  ggtitle("Monday") +
  guides(color=F)
p2 <- ggplot(data=reduced_dehart[reduced_dehart$dayweek=="Sat",], aes(x=drinks)) +
  geom_histogram(binwidth=1) +
  ggtitle("Saturday") +
  guides(color=F)

grid.arrange(p1, p2, nrow=2, top="Comparing the distribution of drinks on Monday vs. Saturday")

reduced_dehart$dayweek <- droplevels(reduced_dehart$dayweek)

monday <- reduced_dehart$drinks[reduced_dehart$dayweek=="Mon"]
saturday <- reduced_dehart$drinks[reduced_dehart$dayweek=="Sat"]
diff <- monday - saturday

#shapiro.test(diff)
#par(mfrow=c(1,2))
#hist(diff, breaks=20)
#qqnorm(diff, pch=1, frame=F)
#qqline(diff, col="blue", lwd=2)
#t.test(monday, saturday, paired=T, alternative="two.sided")

wilcox_results <- wilcox.test(monday, saturday, paired=T, alternative="two.sided")
df.wilcox <- data.frame(V=wilcox_results$statistic,
                        p.value=wilcox_results$p.value)
kable(df.wilcox, align="cc")

# fit poisson mixed effect model, accounting for person as random effect
```

```r
# no interaction
m.pois2 <- glmer(drinks ~ nrel + (1 | id), family="poisson", data=dehart)
sum.pois2 <- summary(m.pois2)
m.pois3 <- glmer(drinks ~ rosn + (1 | id), family="poisson", data=dehart)
sum.pois3 <- summary(m.pois3)
m.pois4 <- glmer(drinks ~ nrel + rosn + (1 | id), family="poisson", data=dehart)
sum.pois4 <- summary(m.pois4)
# include interaction
m.pois5 <- glmer(drinks ~ nrel*rosn + (1 | id), family="poisson", data=dehart)
sum.pois5 <- summary(m.pois5)
m.pois6 <- glmer(drinks ~ nrel*trait_self_esteem + (1 | id), family="poisson", data=dehart)
sum.pois6 <- summary(m.pois6)

# fit negative binomial mixed effect model, accounting for person as random effect
# no interaction
m.nb2 <- glmer.nb(drinks ~ nrel + (1 | id), data=dehart)
sum.nb2 <- summary(m.nb2)
#m.nb3 <- glmer.nb(drinks ~ rosn + (1 | id), data=dehart)
#sum.nb3 <- summary(m.nb3)
#m.nb4 <- glmer.nb(drinks ~ nrel + rosn + (1 | id), data=dehart)
#sum.nb4 <- summary(m.nb4)
# include interaction
#m.nb5 <- glmer.nb(drinks ~ nrel*rosn + (1 | id), data=dehart)
#sum.nb5 <- summary(m.nb5)

# estimate parameters
lambda.pois2 <- sum.pois2$coefficients[1]
lambda.pois6 <- sum.pois6$coefficients[1]
mean.nb2 <- sum.nb2$coefficients[1]
dispersion.nb2 <- mean.nb2^2/(var(dehart$drinks)-mean.nb2)

# generate poisson/nb from estimated parameters
x <- c(0:12)
y_pois2 <- dpois(x, lambda=lambda.pois2)
y_pois6 <- dpois(x, lambda=lambda.pois6)
y_nb2 <- dnbinom(x, size=dispersion.nb2, mu=mean.nb2)

# plot results
hist(dehart$drinks, freq=F, breaks=21, main="Observed data with estimated Poisson and Negative Binomial
points(x+0.1, y_pois2, col="red", type="h", lwd=5)
points(x+0.2, y_nb2, col="blue", type="h", lwd=5)
legend("topright", title="Estimated distributions", legend=c("Poisson", "NegBin"), col=c("red", "blue")

ggResidpanel::resid_panel(m.pois2)

ggResidpanel::resid_panel(m.pois6)

df.pois <- data.frame(fixed.effect=c("intercept","nrel"),
                      estimate=as.numeric(round(sum.pois2$coefficients[,1],3)),
                      Std.Error=as.numeric(round(sum.pois2$coefficients[,2],3)),
                      z.value=as.numeric(round(sum.pois2$coefficients[,3],3)),
                      p.value=as.numeric(round(sum.pois2$coefficients[,4],3)))
kable(df.pois, align="rcccc")
```

```
df.pois.int <- data.frame(fixed.effect=c("intercept","nrel","rosn.medium-low",
                                        "rosn.medium","rosn.medium-high",
                                        "rosn.high","nrel:rosn.medium-low",
                                        "nrel:rosn.medium","nrel:rosn.medium-high",
                                        "nrel:rosn.high"),
                          estimate=as.numeric(round(sum.pois6$coefficients[,1],3)),
                          Std.Error=as.numeric(round(sum.pois6$coefficients[,2],3)),
                          z.value=as.numeric(round(sum.pois6$coefficients[,3],3)),
                          p.value=as.numeric(round(sum.pois6$coefficients[,4],3)))
kable(df.pois.int, align="rcccc")

# plot results
hist(dehart$drinks, freq=F, breaks=21, main="Observed data with estimated Poisson distributions", xlab=
points(x+0.1, y_pois2, col="red", type="h", lwd=5)
points(x+0.2, y_pois6, col="blue", type="h", lwd=5)
legend("topright", title="Estimated distributions", legend=c("Poisson: no interaction", "Poisson: with 

nrel <- seq(from=0, to=10, by=0.1)
trait_self_esteem <- c("low","medium-low","medium","medium-high","high")
nrel_trait_matrix <- expand.grid(nrel,trait_self_esteem)
colnames(nrel_trait_matrix) <- c("nrel","trait_self_esteem")

pred.pois2 <- predict(m.pois2, newdata=as.data.frame(nrel), re.form=NA)
df.pred.pois2 <- data.frame(drinks=pred.pois2, nrel=nrel)

# plot drinks vs nrel main effect
ggplot() +
  geom_jitter(data=dehart, aes(x=nrel, y=drinks),
              width=0.1, height=0.2, alpha=0.3, size=0.5, color="black") +
  geom_smooth(data=df.pred.pois2, aes(x=nrel, y=drinks), method="lm", color="salmon") +
  geom_jitter(data=df.pred.pois2, aes(x=nrel, y=drinks),
              width=0.1, height=0.2, alpha=1, size=0.5, color="red") +
  labs(title="Predicted drinks vs. negative relationship events",
       y="drinks",
       x="index of negative relationship events (scale: 0-10)")

pred.pois6 <- predict(m.pois6, newdata=as.data.frame(nrel_trait_matrix), re.form=NA)
df.pred.pois6 <- data.frame(drinks=pred.pois6,
                            nrel=nrel_trait_matrix$nrel,
                            trait_self_esteem=nrel_trait_matrix$trait_self_esteem)

# plot drinks vs nrel with rosn interaction effects
ggplot() +
  facet_wrap(~ trait_self_esteem) +
  geom_jitter(data=dehart, aes(x=nrel, y=drinks),
              width=0.2, height=0.1, alpha=0.3, size=0.5, color="black") +
  geom_smooth(data=df.pred.pois6, aes(x=nrel, y=drinks, color=trait_self_esteem),
              method="lm") +
  geom_jitter(data=df.pred.pois6, aes(x=nrel, y=drinks, color=trait_self_esteem),
              width=0.2, height=0.1, alpha=1, size=0.5) +
  labs(title="Predicted drinks vs. negative relationship events, faceted by trait self-esteem",
       y="drinks",
       x="index of negative relationship events (scale: 0-10)",
```

```
      color= "trait self-esteem") +
  scale_color_discrete(labels=c("low","medium-low","medium","medium-high","high"))
```