

Homework_4

Lindley Slipetz

7/13/2021

For this homework, I will again be using the childhood adversity and traumatic stress among inpatients at a psychiatric hospital in the Baltimore area from 1993-1995. The data include diagnoses, psychological symptoms, physical and sexual abuse, post-traumatic stress disorder, self-destructive behavior, and demographic data. Since this is a multinomial model rather than an ordered model, I will be using different variables. I will be predicting diagnosis (affective disorders, dissociative disorders, psychotic disorders, and substance use disorders) from positive affect, anxiety, depression, paranoid ideation, psychoticism, sexual impulsiveness, and self harm.

Let's load the packages and data!

```
require(tidyverse)
require(aod) # for wald.test() function
require(dplyr)
require(ggplot2)
require(haven)
require(lmtest) # for lrtest() function
require(MASS)
require(mlogit)
require(MNP)
require(nnet)
full_data <- read.table(file = 'G:\\My Drive\\ICPSR\\ML\\HW_2\\36168-0001-Data.tsv', sep = '\\t', header = TRUE)
```

The diagnosis variable has a ton of levels, so I think it'd be hard to analyze it for specific diagnoses. I'm going to group different diagnoses together into their diagnostic categories so we're left with 4 levels. I will also be eliminating the levels that have very few respondents.

Affective disorders = 0 Dissociative disorders = 1 Psychotic disorders = 2 Substance use disorders = 3

```
full_data <- full_data %>%
  mutate(dx_cat = case_when (
    DX == "AFF" ~ 0,
    DX == "BIP" ~ 0,
    DX == "ANX" ~ 0,
    DX == "DEP" ~ 0,
    DX == "DIS" ~ 1,
    DX == "MPD" ~ 1,
    DX == "SCZ" ~ 2,
    DX == "SUB" ~ 3,
    TRUE ~ as.numeric(NA)))
```

Okay, now let's subset the data.

```
subset_data <- full_data %>%
  dplyr::select(dx_cat, ID, PASUM, SCL_ANX, SCL_DEP, SCL_PAR, SCL_PSY, SISDB_SEXIMP, SISDB_SHARM)
```

Now we're going to deal with missing data.

```
df <- as.data.frame(
  cbind(
    lapply(
      lapply(subset_data, is.na), sum)
    )
  )
```

```
rownames(subset(df, df$V1 != 0))
```

```
## [1] "dx_cat"      "PASUM"      "SCL_ANX"    "SCL_DEP"    "SCL_PAR"
## [6] "SCL_PSY"    "SISDB_SEXIMP" "SISDB_SHARM"
```

Well, there's a lot of columns with missing data (of course we expect dx_cat to have some because we excluded some diagnoses).

Let's count the NA's.

```
sapply(subset_data, function(x) sum(is.na(x)))
```

```
##      dx_cat      ID      PASUM      SCL_ANX      SCL_DEP      SCL_PAR
##      42       0       2       1       1       1
##      SCL_PSY SISDB_SEXIMP SISDB_SHARM
##      1       11       10
```

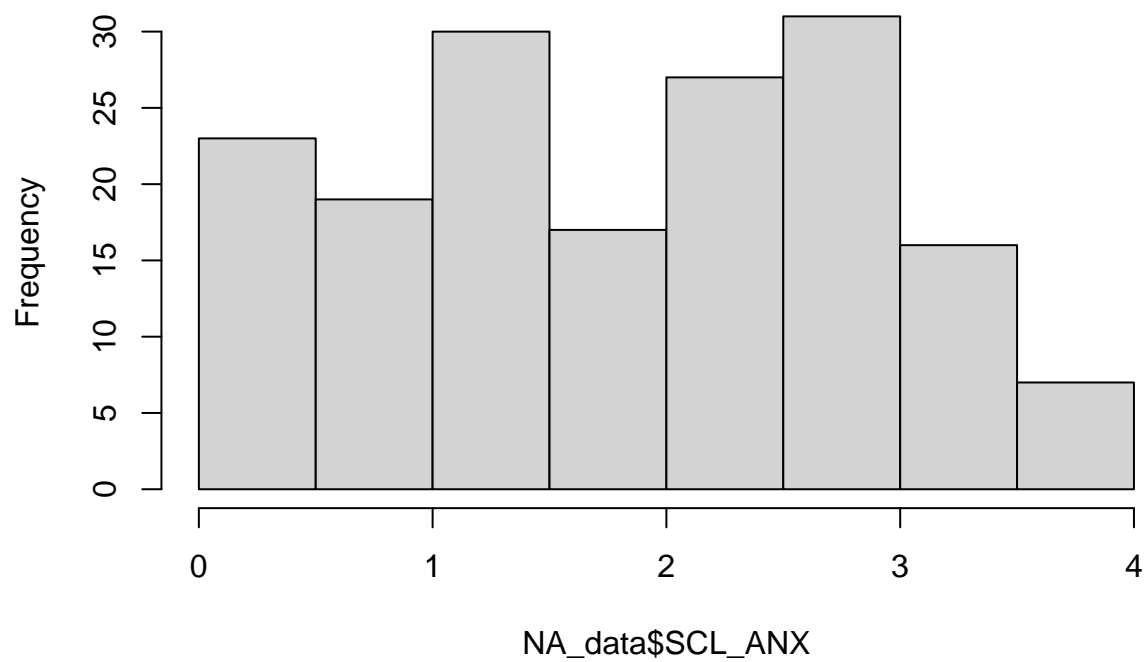
I'm not surprised by the dx_cat missing data and the other variables don't look too bad. I'll just omit the NAs.

```
NA_data <- na.omit(subset_data)
```

Let's look at histograms of the data.

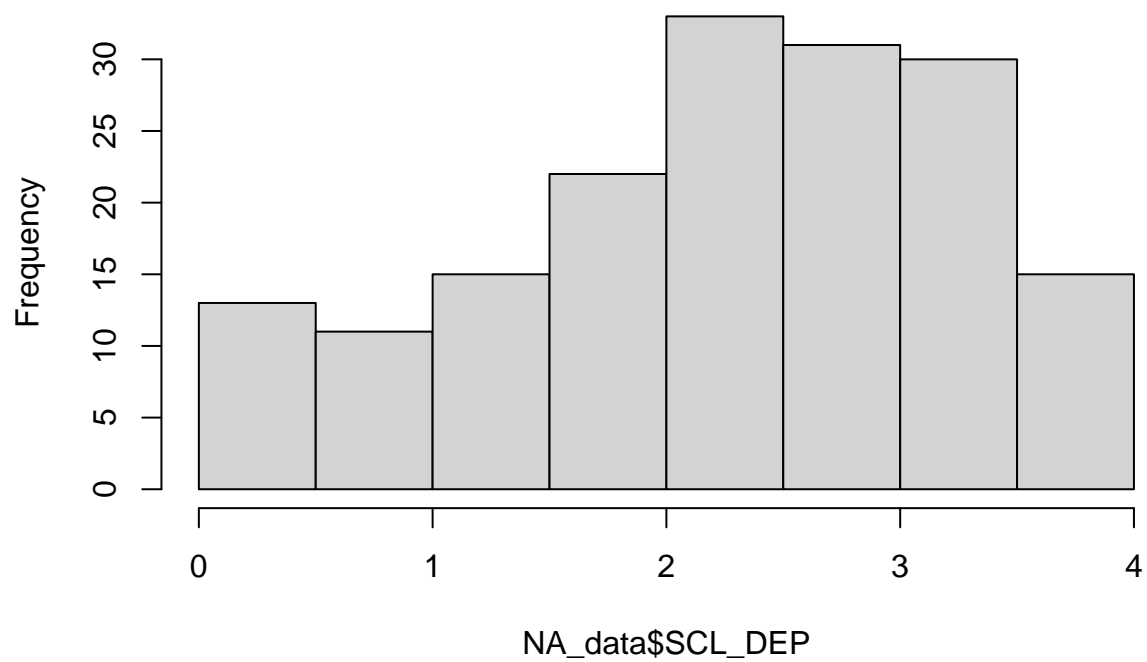
```
hist(NA_data$SCL_ANX)
```

Histogram of NA_data\$SCL_ANX



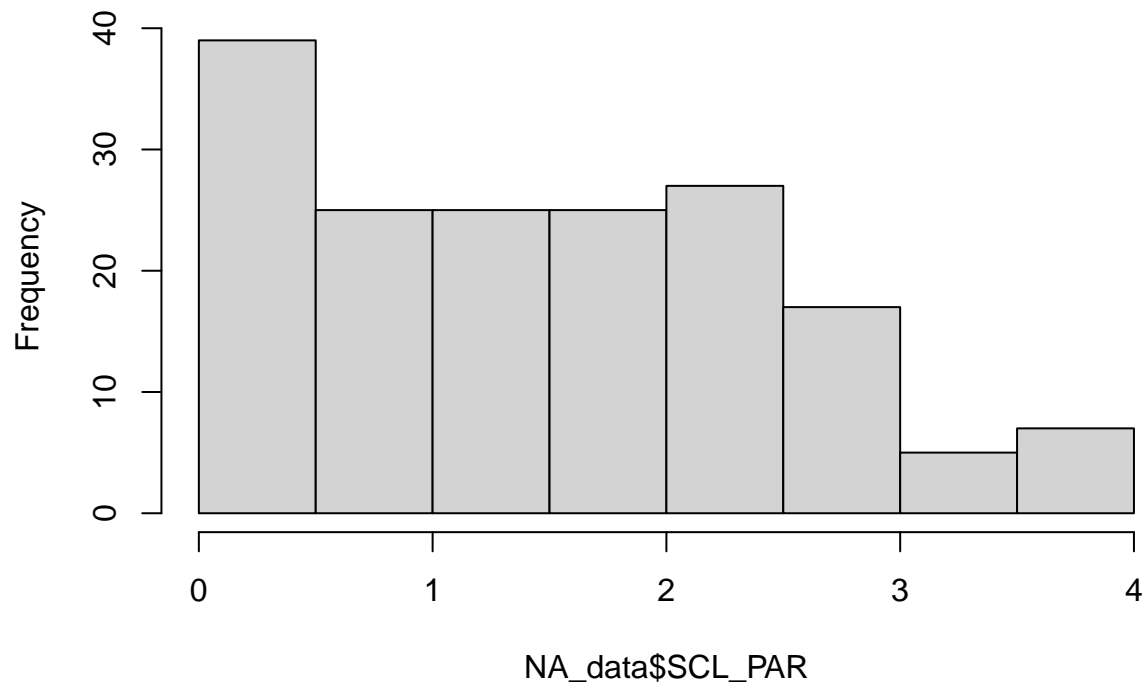
```
hist(NA_data$SCL_DEP)
```

Histogram of NA_data\$SCL_DEP



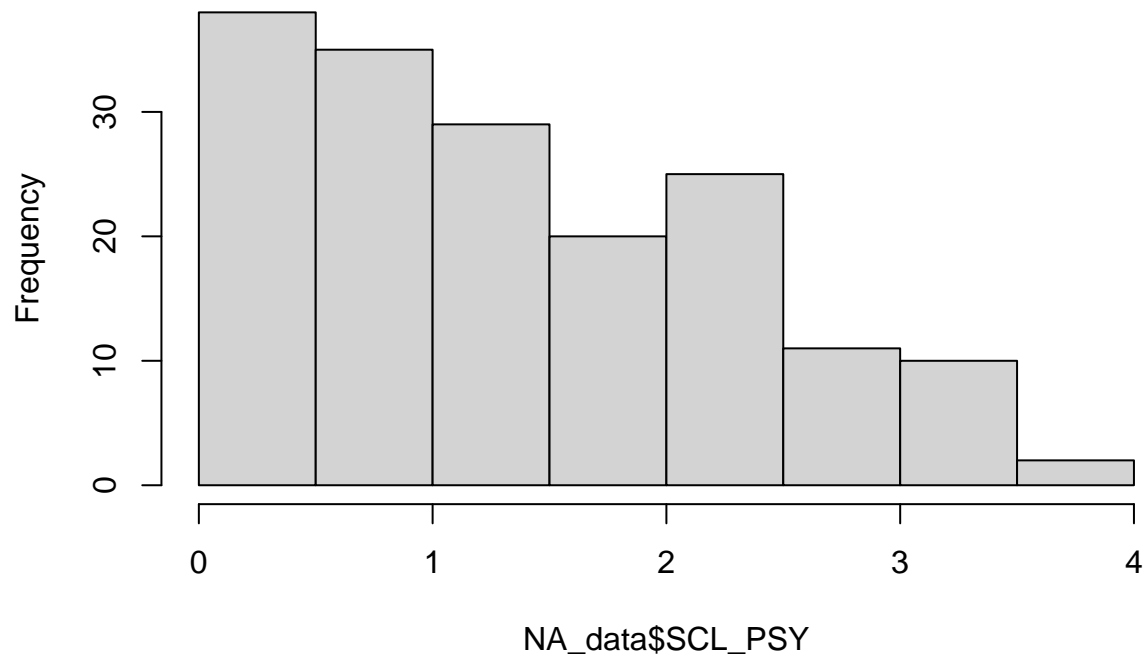
```
hist(NA_data$SCL_PAR)
```

Histogram of NA_data\$SCL_PAR



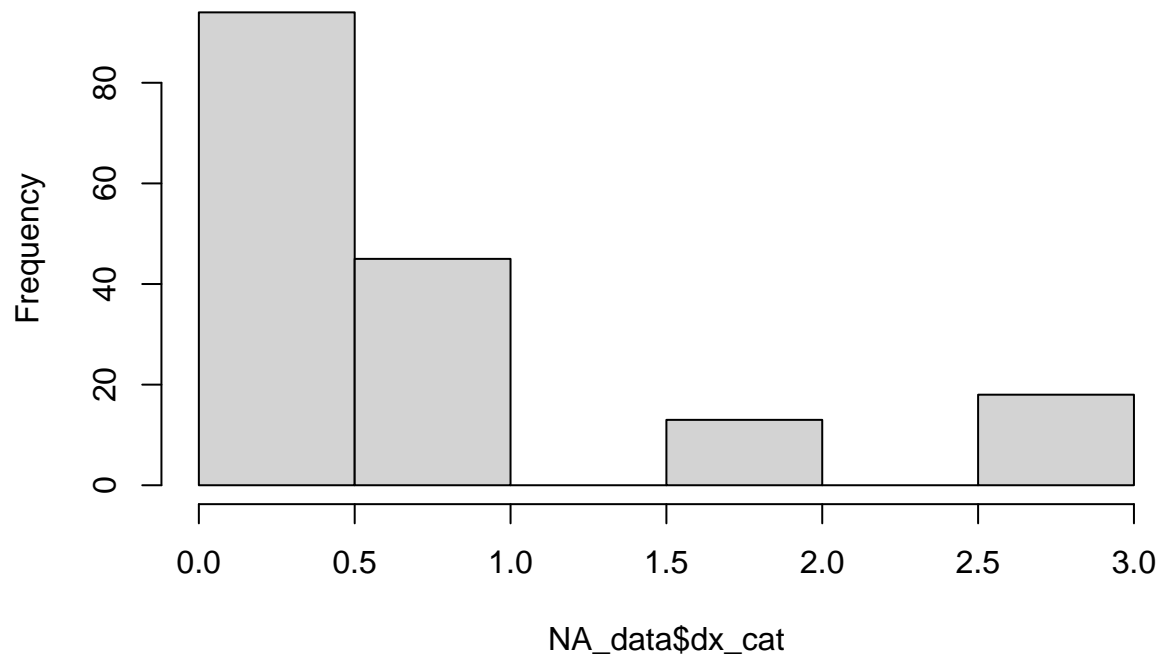
```
hist(NA_data$SCL_PSY)
```

Histogram of NA_data\$SCL_PSY



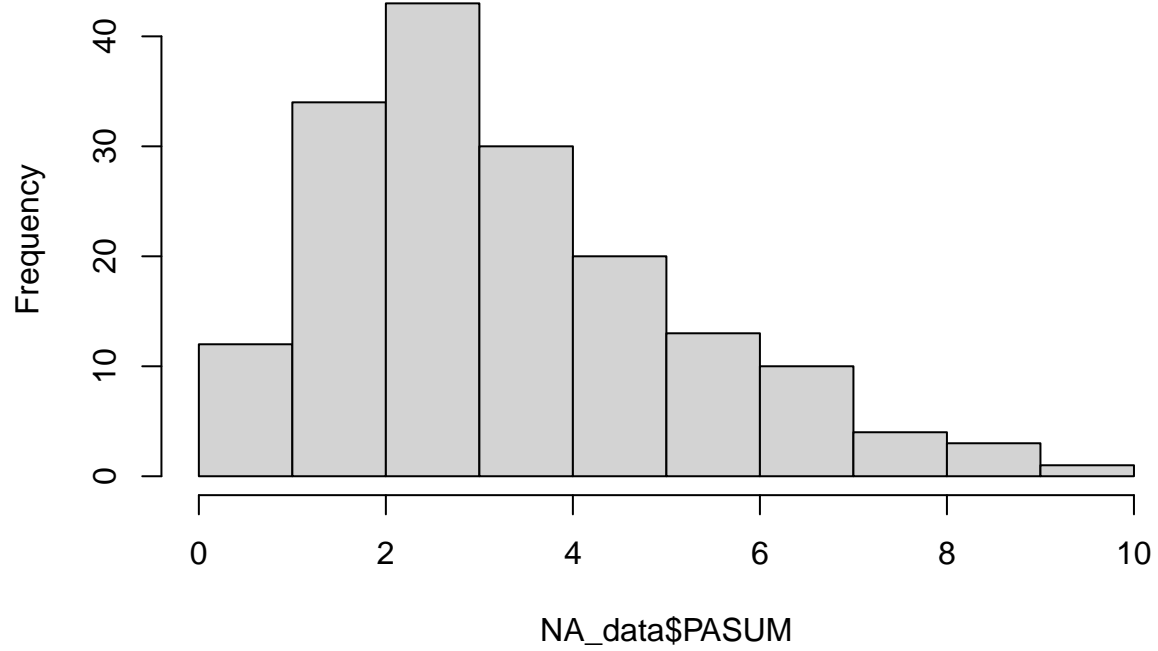
```
hist(NA_data$dx_cat)
```

Histogram of NA_data\$dx_cat



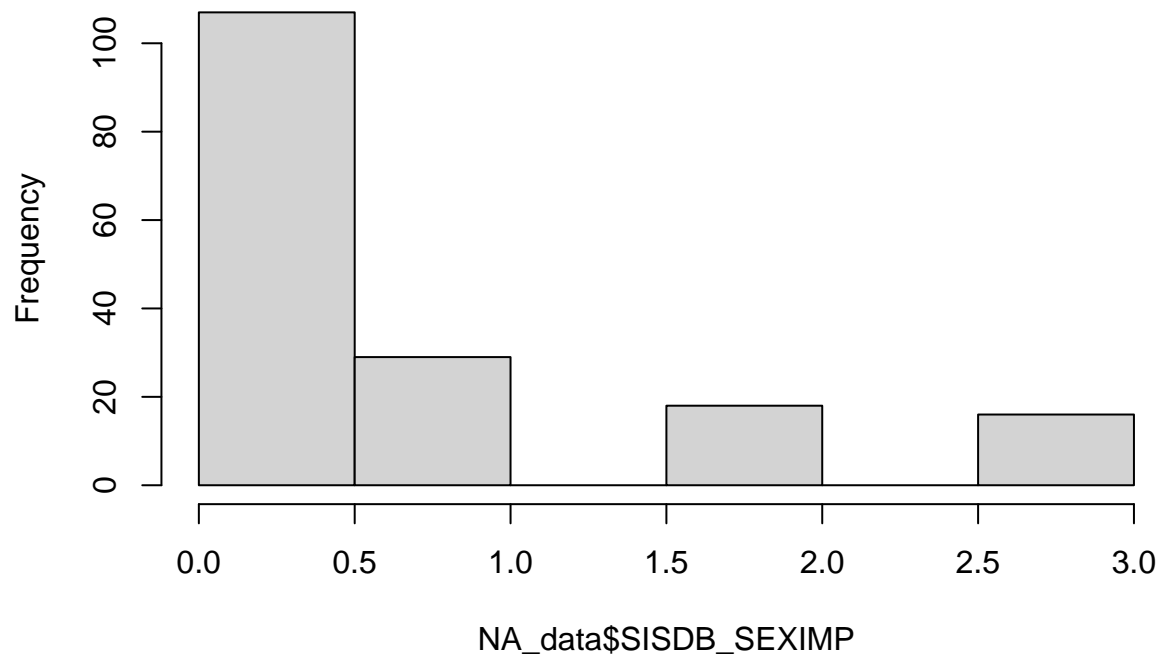
```
hist(NA_data$PASUM)
```

Histogram of NA_data\$PASUM

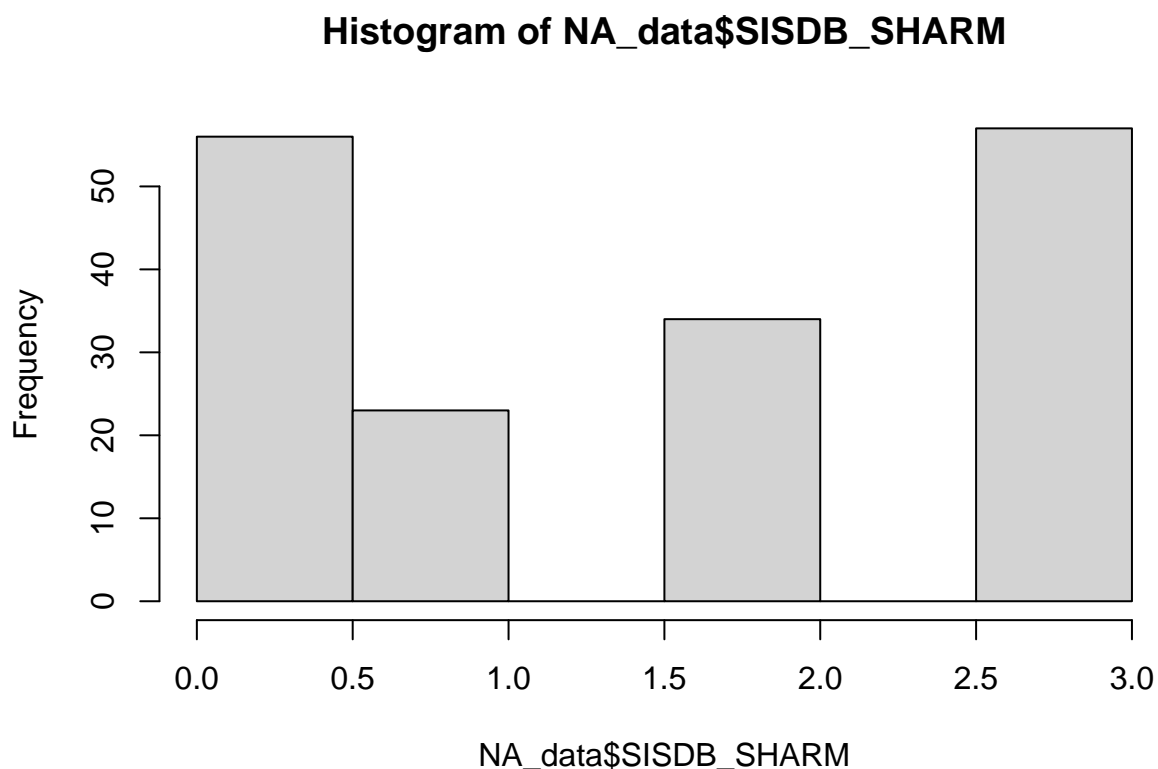


```
hist(NA_data$SISDB_SEXIMP)
```


Histogram of NA_data\$SISDB_SEXIMP



```
hist(NA_data$SISDB_SHARM)
```



Time for a multinomial model!

```
dx_mlogit <- NA_data %>% dplyr::select(ID, dx_cat, PASUM, SCL_ANX, SCL_DEP, SCL_PAR, SCL_PSY, SISDB_SEXIMP, SISDB_SHARM)
dx_mlogit_long <- dfidx(dx_mlogit, choice = "dx_cat", shape = "wide")
head(dx_mlogit_long, n = 12) # Take a look at the data
```

```
## ~~~~~
## first 12 observations out of 680
## ~~~~~
##      ID dx_cat PASUM SCL_ANX SCL_DEP  SCL_PAR SCL_PSY SISDB_SEXIMP SISDB_SHARM
## 1    6 FALSE    7    2.6    2.17 1.333333    0.0          0          0
## 2    6 FALSE    7    2.6    2.17 1.333333    0.0          0          0
## 3    6 FALSE    7    2.6    2.17 1.333333    0.0          0          0
## 4    6  TRUE    7    2.6    2.17 1.333333    0.0          0          0
## 5   11  TRUE    8    2.6    3.23 2.500000    1.5          0          0
## 6   11 FALSE    8    2.6    3.23 2.500000    1.5          0          0
## 7   11 FALSE    8    2.6    3.23 2.500000    1.5          0          0
## 8   11 FALSE    8    2.6    3.23 2.500000    1.5          0          0
## 9   12  TRUE    7    0.7    0.38 0.000000    0.1          0          0
## 10  12 FALSE    7    0.7    0.38 0.000000    0.1          0          0
## 11  12 FALSE    7    0.7    0.38 0.000000    0.1          0          0
## 12  12 FALSE    7    0.7    0.38 0.000000    0.1          0          0
##      idx
## 1  2:0
## 2  2:1
## 3  2:2
## 4  2:3
```

```

## 5 3:0
## 6 3:1
## 7 3:2
## 8 3:3
## 9 4:0
## 10 4:1
## 11 4:2
## 12 4:3
##
## ~~~ indexes ~~~~
##      id1 id2
## 1      2  0
## 2      2  1
## 3      2  2
## 4      2  3
## 5      3  0
## 6      3  1
## 7      3  2
## 8      3  3
## 9      4  0
## 10     4  1
## 11     4  2
## 12     4  3
## indexes: 1, 2
outl_2 <- mlogit(dx_cat ~ 0 | PASUM + SCL_ANX + SCL_DEP + SCL_PAR + SCL_PSY + SISDB_SEXIMP + SISDB_SHARM,
  data = dx_mlogit_long, reflevel = 1)
summary(outl_2)

##
## Call:
## mlogit(formula = dx_cat ~ 0 | PASUM + SCL_ANX + SCL_DEP + SCL_PAR +
##       SCL_PSY + SISDB_SEXIMP + SISDB_SHARM, data = dx_mlogit_long,
##       reflevel = 1, method = "nr")
##
## Frequencies of alternatives:choice
##      0      1      2      3
## 0.552941 0.264706 0.076471 0.105882
##
## nr method
## 6 iterations, 0h:0m:0s
## g'(-H)^-1g = 4.17E-05
## successive function values within tolerance limits
##
## Coefficients :
##              Estimate Std. Error z-value Pr(>|z|)
## (Intercept):1 -1.6005770  0.9008709 -1.7767  0.07562 .
## (Intercept):2 -2.6009377  1.3515565 -1.9244  0.05430 .
## (Intercept):3 -1.0450517  1.0465709 -0.9985  0.31801
## PASUM:1      -0.1123136  0.1280840 -0.8769  0.38055
## PASUM:2       0.1963928  0.1744080  1.1261  0.26014
## PASUM:3      -0.0023038  0.1482323 -0.0155  0.98760
## SCL_ANX:1     0.4614545  0.3357328  1.3745  0.16930
## SCL_ANX:2    -0.9280591  0.6471282 -1.4341  0.15154
## SCL_ANX:3     0.6701386  0.5185853  1.2922  0.19627

```

```
## SCL_DEP:1      -0.1855162  0.3256991 -0.5696  0.56895
## SCL_DEP:2      -0.5989163  0.5378560 -1.1135  0.26548
## SCL_DEP:3      -0.8378238  0.4993626 -1.6778  0.09339 .
## SCL_PAR:1      -0.1083291  0.2721114 -0.3981  0.69055
## SCL_PAR:2       0.6245961  0.4842942  1.2897  0.19715
## SCL_PAR:3       0.0457681  0.3962535  0.1155  0.90805
## SCL_PSY:1       0.1235128  0.3406261  0.3626  0.71690
## SCL_PSY:2       0.9864296  0.6174080  1.5977  0.11011
## SCL_PSY:3       0.2450913  0.4882248  0.5020  0.61566
## SISDB_SEXIMP:1  0.0984264  0.1862122  0.5286  0.59710
## SISDB_SEXIMP:2 -0.2965130  0.4376977 -0.6774  0.49813
## SISDB_SEXIMP:3  0.0943178  0.2594686  0.3635  0.71623
## SISDB_SHARM:1   0.3754810  0.1819697  2.0634  0.03907 *
## SISDB_SHARM:2   0.1460092  0.3162583  0.4617  0.64431
## SISDB_SHARM:3  -0.3958319  0.2578388 -1.5352  0.12474
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -168.18
## McFadden R^2:  0.11178
## Likelihood ratio test : chisq = 42.331 (p.value = 0.0038221)
```

Here's a table of the results.

```
library(stargazer)
```

```
##
## Please cite as:
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
stargazer(out1_2)
```

```
% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Wed, Jul 14, 2021 - 5:37:25 PM
```

Let's look at the average predicted probabilities.

```
colMeans(fitted(out1_2, type = "probabilities"))
```

```
##           0           1           2           3
## 0.55294118 0.26470588 0.07647059 0.10588235
```

The average predicted probability is highest for affective disorders and lowest for psychotic disorders.

Simulating the coefficients.

```
set.seed(1234)
simb <- mvrnorm(n = 1000, mu = coef(out1_2), Sigma = vcov(out1_2))
simb1 <- simb[, seq(1, 16, 2)] # simulated coefficients for dissociative
simb2 <- simb[, seq(2, 17, 2)] # simulated coefficients for psychotic
simb3 <- simb[, seq(3, 18, 2)] # simulated coefficients for substance use
```

Predicted probabilities as function of self-harm.

```
plot_out_1 <- matrix(NA, nrow = 3, ncol = 12)
colnames(plot_out_1) <- paste0(rep(c("Mean_", "Lower_", "Upper_"), times = 4),
rep(1:4, each = 3))
```

Table 1:

		<i>Dependent variable:</i>
		dx_cat
(Intercept):1		−1.601* (0.901)
(Intercept):2		−2.601* (1.352)
(Intercept):3		−1.045 (1.047)
PASUM:1		−0.112 (0.128)
PASUM:2		0.196 (0.174)
PASUM:3		−0.002 (0.148)
SCL_ANX:1		0.461 (0.336)
SCL_ANX:2		−0.928 (0.647)
SCL_ANX:3		0.670 (0.519)
SCL_DEP:1		−0.186 (0.326)
SCL_DEP:2		−0.599 (0.538)
SCL_DEP:3		−0.838* (0.499)
SCL_PAR:1		−0.108 (0.272)
SCL_PAR:2		0.625 (0.484)
SCL_PAR:3		0.046 (0.396)
SCL_PSY:1		0.124 (0.341)
SCL_PSY:2		0.986 (0.617)
SCL_PSY:3	13	0.245 (0.488)
SISDB_SEXIMP:1		0.098

```

for (i in 1:3){ # for each value of education...
# Create the profile + compute the (simulated) linear predictors
Xb1 <- (simb1[, 1] # intercept
+ simb1[, 2] * i # SH
+ simb1[, 3] * mean(NA_data$PASUM, na.rm = TRUE) # PA
+ simb1[, 4] * mean(NA_data$SCL_ANX, na.rm = TRUE) # Anxiety
+ simb1[, 5] * mean(NA_data$SCL_DEP, na.rm = TRUE) # Depression
+ simb1[, 6] * mean(NA_data$SCL_PAR, na.rm = TRUE) # Paranoia
+ simb1[, 7] * mean(NA_data$SCL_PSY, na.rm = TRUE) # Psychotic
+ simb1[, 8] * mean(NA_data$SISDB_SEXIMP, na.rm = TRUE)) # Sex Impulse

Xb2 <- (simb2[, 1] # intercept
+ simb2[, 2] * i # SH
+ simb2[, 3] * mean(NA_data$PASUM, na.rm = TRUE) # PA
+ simb2[, 4] * mean(NA_data$SCL_ANX, na.rm = TRUE) # Anxiety
+ simb2[, 5] * mean(NA_data$SCL_DEP, na.rm = TRUE) # Depression
+ simb2[, 6] * mean(NA_data$SCL_PAR, na.rm = TRUE) # Paranoia
+ simb2[, 7] * mean(NA_data$SCL_PSY, na.rm = TRUE) # Psychotic
+ simb2[, 8] * mean(NA_data$SISDB_SEXIMP, na.rm = TRUE)) # Sex Impulse
Xb3 <- (simb3[, 1] # intercept
+ simb3[, 2] * i # SH
+ simb3[, 3] * mean(NA_data$PASUM, na.rm = TRUE) # PA
+ simb3[, 4] * mean(NA_data$SCL_ANX, na.rm = TRUE) # Anxiety
+ simb3[, 5] * mean(NA_data$SCL_DEP, na.rm = TRUE) # Depression
+ simb3[, 6] * mean(NA_data$SCL_PAR, na.rm = TRUE) # Paranoia
+ simb3[, 7] * mean(NA_data$SCL_PSY, na.rm = TRUE) # Psychotic
+ simb3[, 8] * mean(NA_data$SISDB_SEXIMP, na.rm = TRUE)) # Sex Impulse
denominator <- (1 + exp(Xb2) + exp(Xb3) + exp(Xb1))
probs <- matrix(NA, nrow = 1000, ncol = 4) # matrix to store simulated predicted probs
probs[, 1] <- 1 / denominator
probs[, 2] <- exp(Xb1) / denominator
probs[, 3] <- exp(Xb2) / denominator
probs[, 4] <- exp(Xb3) / denominator
# Store the simulated means and CIs
plot_out_1[i, c(1, 4, 7, 10)] <- colMeans(probs)
CIs <- apply(probs, 2, quantile, probs = c(0.025, 0.975))
plot_out_1[i, 2:3] <- CIs[, 1]
plot_out_1[i, 5:6] <- CIs[, 2]
plot_out_1[i, 8:9] <- CIs[, 3]
plot_out_1[i, 11:12] <- CIs[, 4]
}
plot_out_1

```

##	Mean_1	Lower_1	Upper_1	Mean_2	Lower_2	Upper_2
## [1,]	0.2741534	0.002969966	0.8647190	0.12611213	4.246430e-03	0.5206086
## [2,]	0.2695072	0.002289989	0.8701286	0.08061882	4.002324e-04	0.5428639
## [3,]	0.2541700	0.001709641	0.8704506	0.07407885	2.305142e-05	0.6840438
##	Mean_3	Lower_3	Upper_3	Mean_4	Lower_4	Upper_4
## [1,]	0.02161838	3.515340e-07	0.2038639	0.5781161	0.03215735	0.9693441
## [2,]	0.02021632	2.371278e-07	0.2047829	0.6296576	0.03435427	0.9835028
## [3,]	0.01859379	1.271912e-07	0.1956168	0.6531573	0.03848244	0.9907047

Predicted probabilities of positive affect

```

plot_out_2 <- matrix(NA, nrow = 3, ncol = 12)
colnames(plot_out_2) <- paste0(rep(c("Mean_", "Lower_", "Upper_"), times = 4),
rep(1:4, each = 3))
for (i in 1:3){ # for each value of education...
# Create the profile + compute the (simulated) linear predictors
Xb1 <- (simb1[, 1] # intercept
+ simb1[, 2] * i # PA
+ simb1[, 3] * mean(NA_data$SISDB_SHARM, na.rm = TRUE) # SH
+ simb1[, 4] * mean(NA_data$SCL_ANX, na.rm = TRUE) # Anxiety
+ simb1[, 5] * mean(NA_data$SCL_DEP, na.rm = TRUE) # Depression
+ simb1[, 6] * mean(NA_data$SCL_PAR, na.rm = TRUE) # Paranoia
+ simb1[, 7] * mean(NA_data$SCL_PSY, na.rm = TRUE) # Psychotic
+ simb1[, 8] * mean(NA_data$SISDB_SEXIMP, na.rm = TRUE)) # Sex Impulse

Xb2 <- (simb2[, 1] # intercept
+ simb2[, 2] * i # PA
+ simb2[, 3] * mean(NA_data$SISDB_SHARM, na.rm = TRUE) # SH
+ simb2[, 4] * mean(NA_data$SCL_ANX, na.rm = TRUE) # Anxiety
+ simb2[, 5] * mean(NA_data$SCL_DEP, na.rm = TRUE) # Depression
+ simb2[, 6] * mean(NA_data$SCL_PAR, na.rm = TRUE) # Paranoia
+ simb2[, 7] * mean(NA_data$SCL_PSY, na.rm = TRUE) # Psychotic
+ simb2[, 8] * mean(NA_data$SISDB_SEXIMP, na.rm = TRUE)) # Sex Impulse
Xb3 <- (simb3[, 1] # intercept
+ simb3[, 3] * mean(NA_data$SISDB_SHARM, na.rm = TRUE) # PA
+ simb3[, 2] * i # SH
+ simb3[, 4] * mean(NA_data$SCL_ANX, na.rm = TRUE) # Anxiety
+ simb3[, 5] * mean(NA_data$SCL_DEP, na.rm = TRUE) # Depression
+ simb3[, 6] * mean(NA_data$SCL_PAR, na.rm = TRUE) # Paranoia
+ simb3[, 7] * mean(NA_data$SCL_PSY, na.rm = TRUE) # Psychotic
+ simb3[, 8] * mean(NA_data$SISDB_SEXIMP, na.rm = TRUE)) # Sex Impulse
denominator <- (1 + exp(Xb2) + exp(Xb3) + exp(Xb1))
probs <- matrix(NA, nrow = 1000, ncol = 4) # matrix to store simulated predicted probs
probs[, 1] <- 1 / denominator
probs[, 2] <- exp(Xb1) / denominator
probs[, 3] <- exp(Xb2) / denominator
probs[, 4] <- exp(Xb3) / denominator
# Store the simulated means and CIs
plot_out_2[i, c(1, 4, 7, 10)] <- colMeans(probs)
CIs <- apply(probs, 2, quantile, probs = c(0.025, 0.975))
plot_out_2[i, 2:3] <- CIs[, 1]
plot_out_2[i, 5:6] <- CIs[, 2]
plot_out_2[i, 8:9] <- CIs[, 3]
plot_out_2[i, 11:12] <- CIs[, 4]
}
plot_out_2

```

```

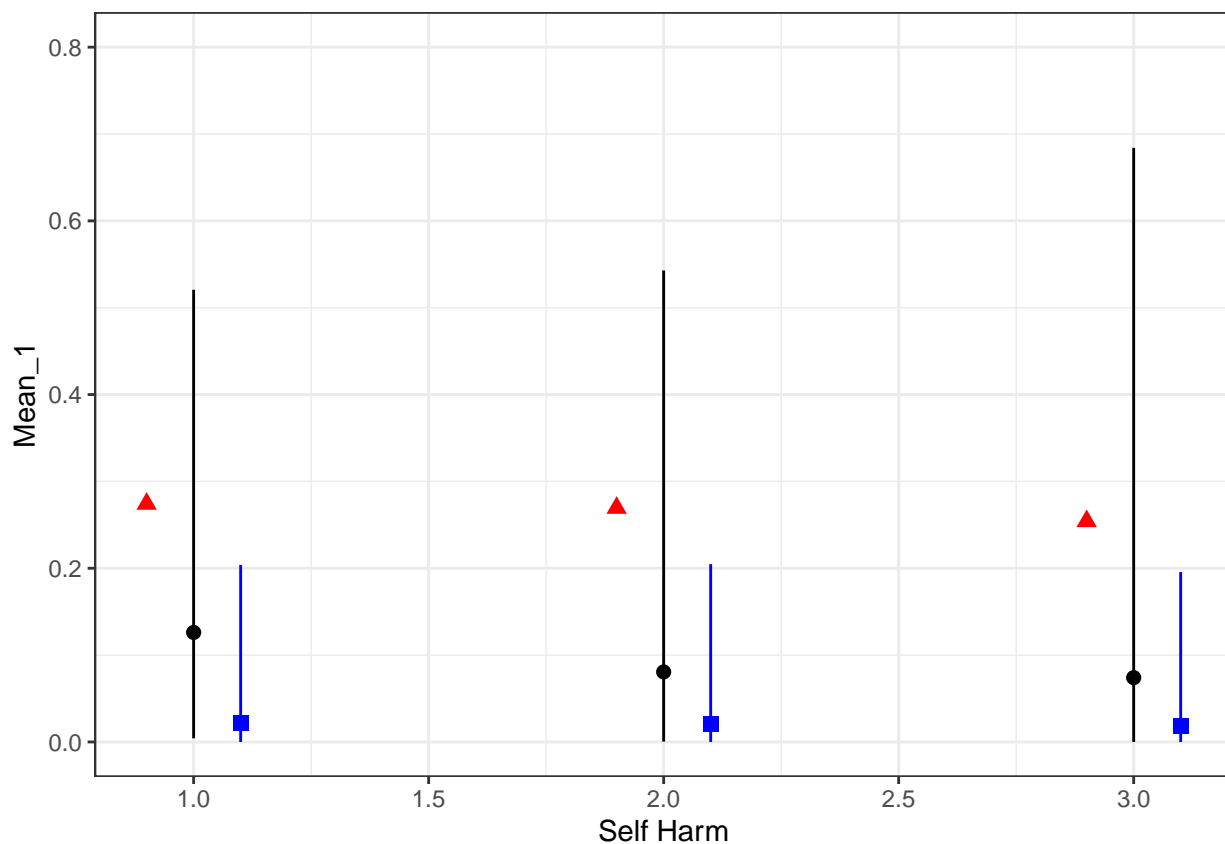
##           Mean_1      Lower_1      Upper_1      Mean_2      Lower_2      Upper_2
## [1,] 0.3842268 0.010745011 0.8974508 0.16193536 3.946234e-03 0.6102198
## [2,] 0.3815008 0.008765205 0.9109551 0.10856730 4.172799e-04 0.6318073
## [3,] 0.3620436 0.006095583 0.9122809 0.09628875 1.874315e-05 0.7332028
##           Mean_3      Lower_3      Upper_3      Mean_4      Lower_4      Upper_4
## [1,] 0.02306093 1.158041e-06 0.2089118 0.4307769 0.02134511 0.8801300
## [2,] 0.02169836 8.300157e-07 0.2069103 0.4882336 0.02352418 0.9382509
## [3,] 0.02001922 4.089577e-07 0.1984033 0.5216484 0.02540563 0.9635710

```

Plot of self harm.

```
plot_out_1 <- as.data.frame(plot_out_1)
plot_out_1$SISDB_SHARM <- c(1:3)
g1 <- ggplot(data = plot_out_1) +
  geom_pointrange(aes(x = SISDB_SHARM - 0.1, y = Mean_1, ymin = Lower_1, ymax = Upper_1),
    shape = 17, color = "red") +
  geom_pointrange(aes(x = SISDB_SHARM, y = Mean_2, ymin = Lower_2, ymax = Upper_2),
    shape = 16, color = "black") +
  geom_pointrange(aes(x = SISDB_SHARM + 0.1, y = Mean_3, ymin = Lower_3, ymax = Upper_3),
    shape = 15, color = "blue") +
  xlab("Self Harm") +
  ylim(0.0, 0.8) + theme_bw()
g1
```

Warning: Removed 3 rows containing missing values (geom_segment).



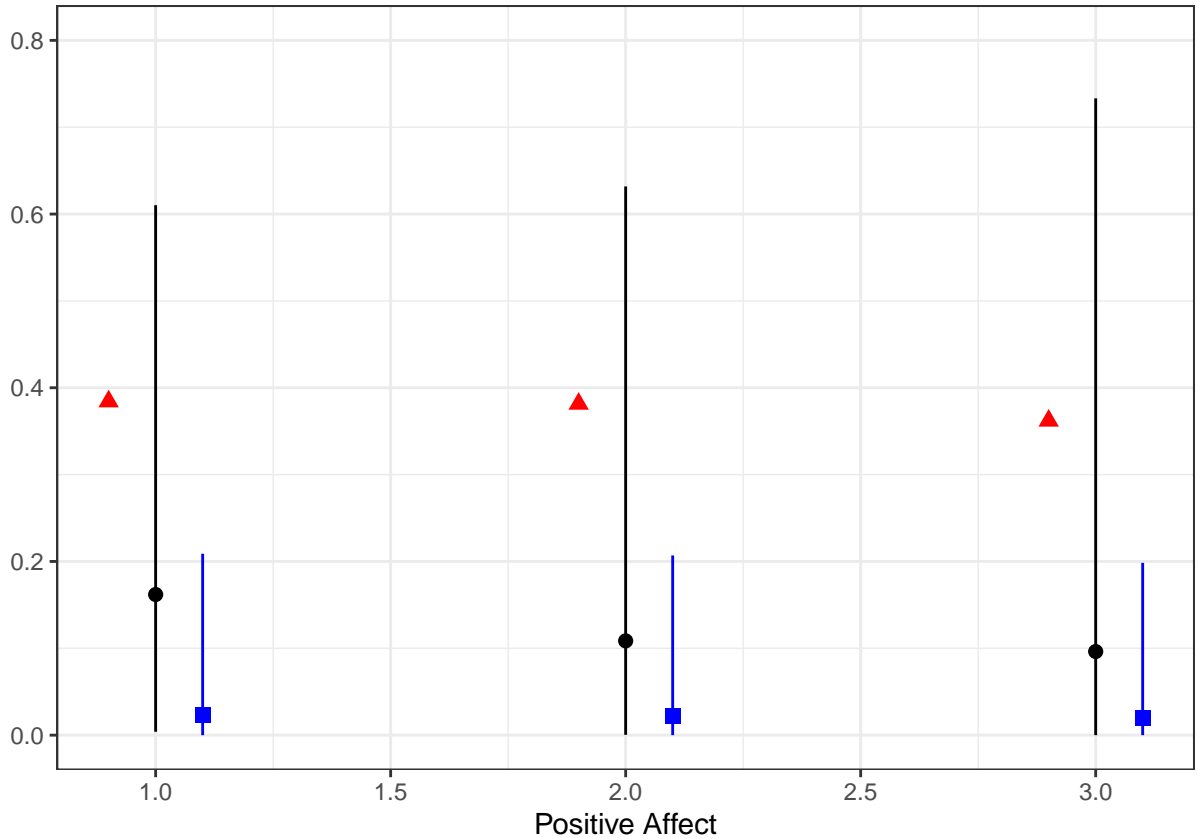
The graph shows there's not much of a difference across levels of self harm.

```
plot_out_2 <- as.data.frame(plot_out_2)
plot_out_2$PASUM <- c(1:3)
g2 <- ggplot(data = plot_out_2) +
  geom_pointrange(aes(x = PASUM - 0.1, y = Mean_1, ymin = Lower_1, ymax = Upper_1),
    shape = 17, color = "red") +
  geom_pointrange(aes(x = PASUM, y = Mean_2, ymin = Lower_2, ymax = Upper_2),
    shape = 16, color = "black") +
  geom_pointrange(aes(x = PASUM + 0.1, y = Mean_3, ymin = Lower_3, ymax = Upper_3),
    shape = 15, color = "blue") +
```



```
xlab("Positive Affect") + ylab("") +
ylim(0.0, 0.8) + theme_bw()
g2
```

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



Psychotic disorders show some difference between low positive affect and mid to high positive affect, but the actual coefficient is not significant.

Let's try a different baseline.

```
out1_1 <- mlogit(dx_cat ~ 0 | PASUM + SCL_ANX + SCL_DEP + SCL_PAR + SCL_PSY + SISDB_SEXIMP + SISDB_SHARM,
  data = dx_mlogit_long, reflevel = 2)
summary(out1_1)
```

```
##
## Call:
## mlogit(formula = dx_cat ~ 0 | PASUM + SCL_ANX + SCL_DEP + SCL_PAR +
##       SCL_PSY + SISDB_SEXIMP + SISDB_SHARM, data = dx_mlogit_long,
##       reflevel = 2, method = "nr")
##
## Frequencies of alternatives:choice
##      1      0      2      3
## 0.264706 0.552941 0.076471 0.105882
##
## nr method
## 6 iterations, 0h:0m:0s
```

```
## g'(-H)^-1g = 4.17E-05
## successive function values within tolerance limits
##
## Coefficients :
##           Estimate Std. Error z-value Pr(>|z|)
## (Intercept):0    1.6005770   0.9008709   1.7767 0.075618 .
## (Intercept):2   -1.0003607   1.5226148  -0.6570 0.511180
## (Intercept):3    0.5555253   1.2510335   0.4441 0.657004
## PASUM:0          0.1123136   0.1280840   0.8769 0.380555
## PASUM:2          0.3087064   0.2013262   1.5334 0.125186
## PASUM:3          0.1100098   0.1770578   0.6213 0.534388
## SCL_ANX:0        -0.4614545   0.3357328  -1.3745 0.169296
## SCL_ANX:2        -1.3895136   0.6917084  -2.0088 0.044557 *
## SCL_ANX:3         0.2086841   0.5599978   0.3727 0.709408
## SCL_DEP:0         0.1855162   0.3256991   0.5696 0.568953
## SCL_DEP:2        -0.4134001   0.5824858  -0.7097 0.477880
## SCL_DEP:3        -0.6523076   0.5426709  -1.2020 0.229351
## SCL_PAR:0         0.1083291   0.2721114   0.3981 0.690552
## SCL_PAR:2         0.7329251   0.5209202   1.4070 0.159433
## SCL_PAR:3         0.1540972   0.4345600   0.3546 0.722885
## SCL_PSY:0        -0.1235128   0.3406261  -0.3626 0.716900
## SCL_PSY:2         0.8629169   0.6592986   1.3088 0.190588
## SCL_PSY:3         0.1215786   0.5356583   0.2270 0.820447
## SISDB_SEXIMP:0   -0.0984264   0.1862122  -0.5286 0.597103
## SISDB_SEXIMP:2   -0.3949394   0.4515055  -0.8747 0.381728
## SISDB_SEXIMP:3   -0.0041086   0.2827377  -0.0145 0.988406
## SISDB_SHARM:0    -0.3754810   0.1819697  -2.0634 0.039072 *
## SISDB_SHARM:2    -0.2294718   0.3399473  -0.6750 0.499662
## SISDB_SHARM:3    -0.7713129   0.2853869  -2.7027 0.006878 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -168.18
## McFadden R^2:  0.11178
## Likelihood ratio test : chisq = 42.331 (p.value = 0.0038221)
```

The model now has three significant coefficients. This difference makes sense because there is a different baseline, so the coefficients are interpreted relative to a different baseline. We still see the significant self harm coefficient for 1 relative to 0 (which we had for 0 relative to 1).

Now we'll test the IIA assumption. First, estimate models excluding each choice

```
out1 <- mlogit(dx_cat ~ 0 | PASUM + SCL_ANX + SCL_DEP + SCL_PAR + SCL_PSY + SISDB_SEXIMP + SISDB_SHARM,
alt.subset = c("2", "3"))
out2 <- mlogit(dx_cat ~ 0 | PASUM + SCL_ANX + SCL_DEP + SCL_PAR + SCL_PSY + SISDB_SEXIMP + SISDB_SHARM,
alt.subset = c("1", "3"))
out3 <- mlogit(dx_cat ~ 0 | PASUM + SCL_ANX + SCL_DEP + SCL_PAR + SCL_PSY + SISDB_SEXIMP + SISDB_SHARM,
alt.subset = c("1", "2"))
```

Conduct Hausman-McFadden tests

```
hmfptest(out1_1, out1)
```

```
##
## Hausman-McFadden test
##
```

```
## data: dx_mlogit_long
## chisq = -38.934, df = 8, p-value = 1
## alternative hypothesis: IIA is rejected
```

A p-value of 1...well, that's not good. IIA has been rejected.

Predicted probabilities with affective disorders left out.

```
out4 <- mlogit(dx_cat ~ 0 | PASUM + SCL_ANX + SCL_DEP + SCL_PAR + SCL_PSY + SISDB_SEXIMP + SISDB_SHARM,
alt.subset = c("1", "2", "3"))
colMeans(fitted(out4, type = "probabilities"))
```

```
##           1           2           3
## 0.5921053 0.1710526 0.2368421
```

The proportions between the categories are maintained when affective disorders are removed (though the values individually increase).