Visualizations

George Rhodes

5/7/2018

```
## Installing package into '/Users/george/Documents/School/UW/Thesis/thesis_R/packrat/lib/x86_64-apple-d
## (as 'lib' is unspecified)

##
## The downloaded binary packages are in
##  /var/folders/6s/cv2xpvws1978z4cyr4w1sl0h0000gn/T//RtmpfbpWM5/downloaded_packages

## Warning: package 'haven' was built under R version 3.4.3

## Installing package into '/Users/george/Documents/School/UW/Thesis/thesis_R/packrat/lib/x86_64-apple-d
## (as 'lib' is unspecified)

##
## The downloaded binary packages are in
##  /var/folders/6s/cv2xpvws1978z4cyr4w1sl0h0000gn/T//RtmpfbpWM5/downloaded_packages

## Warning: package 'dplyr' was built under R version 3.4.2

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## Installing package into '/Users/george/Documents/School/UW/Thesis/thesis_R/packrat/lib/x86_64-apple-d
## (as 'lib' is unspecified)

##
## The downloaded binary packages are in
##  /var/folders/6s/cv2xpvws1978z4cyr4w1sl0h0000gn/T//RtmpfbpWM5/downloaded_packages

## Installing package into '/Users/george/Documents/School/UW/Thesis/thesis_R/packrat/lib/x86_64-apple-d
## (as 'lib' is unspecified)

##
##   There is a binary version available (and will be installed) but
##   the source version is later:
##      binary source
## MASS 7.3-49 7.3-50
##
##
## The downloaded binary packages are in
##  /var/folders/6s/cv2xpvws1978z4cyr4w1sl0h0000gn/T//RtmpfbpWM5/downloaded_packages

## Warning: package 'MASS' was built under R version 3.4.3

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

```
## The following object is masked from 'package:dplyr':
##
##     select
## Installing package into '/Users/george/Documents/School/UW/Thesis/thesis_R/packrat/lib/x86_64-apple-d
## (as 'lib' is unspecified)
##
## The downloaded binary packages are in
##  /var/folders/6s/cv2xpvws1978z4cyr4w1sl0h0000gn/T//RtmpfbpWM5/downloaded_packages
## Warning: package 'pscl' was built under R version 3.4.2
## Classes and Methods for R developed in the
## Political Science Computational Laboratory
## Department of Political Science
## Stanford University
## Simon Jackman
## hurdle and zeroinfl functions by Achim Zeileis
## Installing package into '/Users/george/Documents/School/UW/Thesis/thesis_R/packrat/lib/x86_64-apple-d
## (as 'lib' is unspecified)
##
## The downloaded binary packages are in
##  /var/folders/6s/cv2xpvws1978z4cyr4w1sl0h0000gn/T//RtmpfbpWM5/downloaded_packages
## Installing package into '/Users/george/Documents/School/UW/Thesis/thesis_R/packrat/lib/x86_64-apple-d
## (as 'lib' is unspecified)
##
## The downloaded binary packages are in
##  /var/folders/6s/cv2xpvws1978z4cyr4w1sl0h0000gn/T//RtmpfbpWM5/downloaded_packages
## Installing package into '/Users/george/Documents/School/UW/Thesis/thesis_R/packrat/lib/x86_64-apple-d
## (as 'lib' is unspecified)
##
## The downloaded binary packages are in
##  /var/folders/6s/cv2xpvws1978z4cyr4w1sl0h0000gn/T//RtmpfbpWM5/downloaded_packages
##
## Attaching package: 'scales'
## The following object is masked from 'package:readr':
##
##     col_factor
## Installing package into '/Users/george/Documents/School/UW/Thesis/thesis_R/packrat/lib/x86_64-apple-d
## (as 'lib' is unspecified)
##
##   There is a binary version available (and will be installed) but
##   the source version is later:
##          binary source
## ggthemes  3.4.2  3.5.0
##
##
## The downloaded binary packages are in
##  /var/folders/6s/cv2xpvws1978z4cyr4w1sl0h0000gn/T//RtmpfbpWM5/downloaded_packages
## Warning: package 'ggthemes' was built under R version 3.4.4
```

```
## Installing package into '/Users/george/Documents/School/UW/Thesis/thesis_R/packrat/lib/x86_64-apple-
## (as 'lib' is unspecified)

##
## The downloaded binary packages are in
##  /var/folders/6s/cv2xpvws1978z4cyr4w1sl0h0000gn/T//RtmpfbpWM5/downloaded_packages

## ------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## Installing package into '/Users/george/Documents/School/UW/Thesis/thesis_R/packrat/lib/x86_64-apple-
## (as 'lib' is unspecified)

##
##   There is a binary version available (and will be installed) but
##   the source version is later:
##          binary source
## ggthemes  3.4.2  3.5.0
##
##
## The downloaded binary packages are in
##  /var/folders/6s/cv2xpvws1978z4cyr4w1sl0h0000gn/T//RtmpfbpWM5/downloaded_packages

## Installing package into '/Users/george/Documents/School/UW/Thesis/thesis_R/packrat/lib/x86_64-apple-
## (as 'lib' is unspecified)

##
## The downloaded binary packages are in
##  /var/folders/6s/cv2xpvws1978z4cyr4w1sl0h0000gn/T//RtmpfbpWM5/downloaded_packages
```

Negative Binomial: Binge 0-30 w/era FAMILY INCOME

```
##
## Call:
## glm.nb(formula = dr5day ~ faminc_mid + treated + alctry + evermj +
##     evercoc + edu4cat + age + sex + marital + dependents + era +
##     faminc_mid * treated + edu4cat * treated, data = data_clean,
##     na.action = na.omit, init.theta = 0.1968806455, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3375  -0.8167  -0.6511  -0.3011   7.5489
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   1.266054   0.059346  21.333  < 2e-16 ***
## faminc_mid                   -0.002023   0.003257  -0.621 0.534483
## treatedDistant                0.237258   0.086056   2.757 0.005833 **
## treatedRecent                 0.351517   0.123686   2.842 0.004483 **
## alctry                       -0.062265   0.002340 -26.609  < 2e-16 ***
## evermjyes                     0.580383   0.019689  29.478  < 2e-16 ***
## evercocyes                    0.493810   0.023155  21.326  < 2e-16 ***
## edu4catHS                    -0.053800   0.029333  -1.834 0.066637 .
## edu4catSome C                -0.339554   0.030856 -11.004  < 2e-16 ***
## edu4catC Grad                -0.592541   0.031961 -18.539  < 2e-16 ***
## age 30-34                    -0.143802   0.027922  -5.150  2.6e-07 ***
## age 35-49                    -0.290318   0.025264 -11.491  < 2e-16 ***
## age 50-64                    -0.548228   0.031686 -17.302  < 2e-16 ***
## age 65+                      -0.982528   0.041003 -23.962  < 2e-16 ***
## sex f                        -0.994521   0.017013 -58.456  < 2e-16 ***
## marital widowed               0.168130   0.049257   3.413 0.000642 ***
## marital divorced              0.373008   0.024696  15.104  < 2e-16 ***
## marital never                 0.264946   0.024732  10.713  < 2e-16 ***
## dependents1                  -0.266323   0.024759 -10.756  < 2e-16 ***
## dependents2                  -0.290200   0.025544 -11.361  < 2e-16 ***
## dependents3+                 -0.351295   0.031521 -11.145  < 2e-16 ***
## era                           0.200419   0.016803  11.928  < 2e-16 ***
## faminc_mid:treatedDistant    -0.014904   0.011308  -1.318 0.187497
## faminc_mid:treatedRecent     -0.020934   0.020624  -1.015 0.310099
## treatedDistant:edu4catHS      0.007599   0.095681   0.079 0.936701
## treatedRecent:edu4catHS      -0.204294   0.149156  -1.370 0.170792
## treatedDistant:edu4catSome C  0.078276   0.101325   0.773 0.439804
## treatedRecent:edu4catSome C   0.135343   0.158985   0.851 0.394607
## treatedDistant:edu4catC Grad  0.355146   0.115749   3.068 0.002153 **
## treatedRecent:edu4catC Grad   0.655067   0.207688   3.154 0.001610 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(0.1969) family taken to be 1)
##
##     Null deviance: 74526  on 99800  degrees of freedom
## Residual deviance: 60665  on 99771  degrees of freedom
## AIC: 226800
##
```

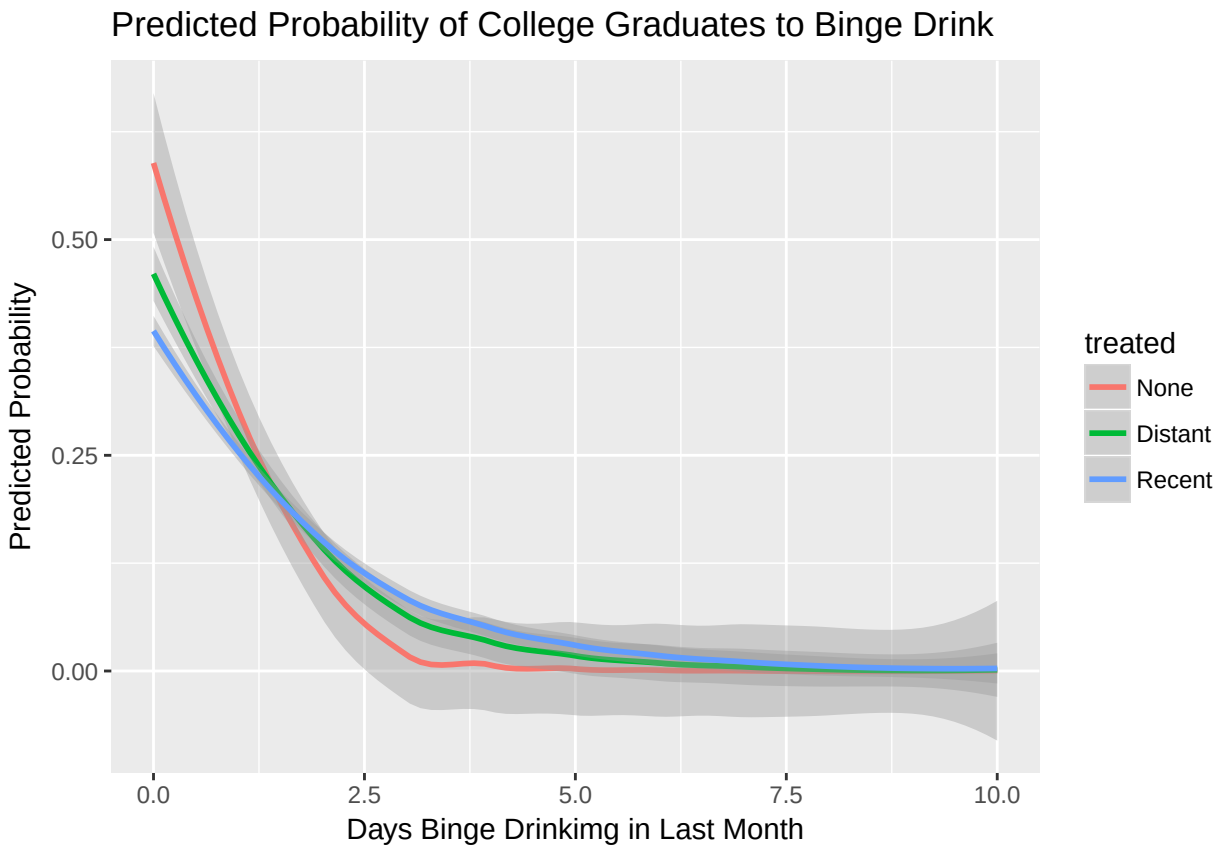
```
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  0.19688
##           Std. Err.:  0.00175
##
##  2 x log-likelihood:  -226738.23000
```

```r
#creating data fixed for visualizations
data_fixed <- data.frame(faminc_mid = mean(data_clean$faminc_mid), treated = 0:2, alctry = mean(data_cl
                         evermj = 2, evercoc = 2, edu4cat = 3, age = 15, sex = 1, marital = 1,
                         dependents = 0, era = 1)
data_fixed_fac <- data_fixed


#convert all to data_fixed_fac factors matching model
data_fixed_fac$treated <- factor(data_fixed$treated, labels =
                                   c("None", "Distant", "Recent"))

data_fixed_fac$evermj <- factor(data_fixed$evermj, labels =
                                  c("yes"))

data_fixed_fac$evercoc <- factor(data_fixed$evercoc, labels =
                                   c("yes"))

data_fixed_fac$edu4cat <- factor(data_fixed$edu4cat, labels = c("C Grad"))

data_fixed_fac$age <-factor(data_fixed$age, labels =
                              c(" 35-49"))

data_fixed_fac$sex <- factor(data_fixed$sex, labels = c(" m"))

data_fixed_fac$marital <-factor(data_fixed$marital, labels =
                                  c(" married"))

data_fixed_fac$dependents <-factor(data_fixed$dependents, labels =
                                     c(" 0"))


#Determine prediction for fixed characteristics and treated (none, distant, recent)
prediction <- predict.glm(nb_faminc, data_fixed_fac)

#determine probability_dnbinom for each prediction
prob_dnbinom <- 1- (prediction/ (prediction + 1))


#run dnbinom for each probability[1:3]
none_dnbinom <- dnbinom(0:10, 1, prob_dnbinom[1])

distant_dnbinom <- dnbinom(0:10, 1, prob_dnbinom[2])

recent_dnbinom <- dnbinom(0:10, 1, prob_dnbinom[3])


#create dataframe for graph for each none/distant/recent_dnbinom
none_data_dnbinom <- data.frame(dr5day = 0:10, dense = none_dnbinom, treated = 0)

distant_data_dnbinom <- data.frame(dr5day = 0:10, dense = distant_dnbinom, treated = 1)

recent_data_dnbinom <- data.frame(dr5day = 0:10, dense = recent_dnbinom, treated = 2)
```

```r
#stack dataframes into data_dnbinom
stack5 <- rbind(none_data_dnbinom, distant_data_dnbinom)
data_dnbinom <-rbind(stack5, recent_data_dnbinom)

#make treated as factors to match model nb_faminc
data_dnbinom$treated <- factor(data_dnbinom$treated, labels =
                                    c("None", "Distant", "Recent"))
```

```r
#plot it
plot_dnbinom <- ggplot(data_dnbinom,
                       mapping = aes(x = dr5day, y = dense, color = treated))+
  geom_smooth()+
  labs(title = "Predicted Probability of College Graduates to Binge Drink", x = "Days Binge Drinkimg in
       y = "Predicted Probability")
plot_dnbinom
```

```
## `geom_smooth()` using method = 'loess'
```



Model Used for FACET: faminc negative binomial with treatment as yes/no

```
##
## Call:
## glm.nb(formula = dr5day ~ faminc_mid + treatedyn + alctry + evermj +
##     evercoc + edu4cat + age + sex + marital + dependents + era +
##     faminc_mid * treatedyn + edu4cat * treatedyn, data = data_clean,
##     na.action = na.omit, init.theta = 0.1968342402, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3425  -0.8168  -0.6511  -0.3014   7.5309
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  1.263300   0.059345  21.288  < 2e-16 ***
## faminc_mid                  -0.001993   0.003257  -0.612 0.540700
## treatedynYes                 0.274860   0.073038   3.763 0.000168 ***
## alctry                      -0.062083   0.002340 -26.535  < 2e-16 ***
## evermjyes                    0.579775   0.019689  29.446  < 2e-16 ***
## evercocyes                   0.494458   0.023152  21.357  < 2e-16 ***
## edu4catHS                   -0.053604   0.029335  -1.827 0.067655 .
## edu4catSome C               -0.339220   0.030858 -10.993  < 2e-16 ***
## edu4catC Grad               -0.592248   0.031963 -18.529  < 2e-16 ***
## age 30-34                   -0.143790   0.027923  -5.149 2.61e-07 ***
## age 35-49                   -0.290641   0.025262 -11.505  < 2e-16 ***
## age 50-64                   -0.548763   0.031677 -17.324  < 2e-16 ***
## age 65+                     -0.983248   0.040997 -23.983  < 2e-16 ***
## sex f                       -0.995666   0.017013 -58.524  < 2e-16 ***
## marital widowed              0.171818   0.049237   3.490 0.000484 ***
## marital divorced             0.373053   0.024695  15.107  < 2e-16 ***
## marital never                0.264968   0.024731  10.714  < 2e-16 ***
## dependents1                 -0.266053   0.024758 -10.746  < 2e-16 ***
## dependents2                 -0.290362   0.025544 -11.367  < 2e-16 ***
## dependents3+                -0.351412   0.031520 -11.149  < 2e-16 ***
## era                          0.200543   0.016803  11.935  < 2e-16 ***
## faminc_mid:treatedynYes     -0.016556   0.010019  -1.652 0.098448 .
## treatedynYes:edu4catHS      -0.048978   0.082461  -0.594 0.552542
## treatedynYes:edu4catSome C   0.084091   0.087470   0.961 0.336367
## treatedynYes:edu4catC Grad   0.405009   0.102117   3.966 7.30e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(0.1968) family taken to be 1)
##
##     Null deviance: 74515  on 99800  degrees of freedom
## Residual deviance: 60664  on 99776  degrees of freedom
## AIC: 226798
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  0.19683
##           Std. Err.:  0.00174
```

```
##
##  2 x log-likelihood:  -226746.15500
```

Facet Graphs

```
#2.0 Compare lines of college grad and <HS, facet on treatedyn
#creating data fixed for visualizations treatedyn yes and no
data_fixed2.0no <- data.frame(faminc_mid = mean(data_clean$faminc_mid),
                         treatedyn = 0, alctry = mean(data_clean$alctry),
                         evermj = 2, evercoc = 2, edu4cat = 1:2, age = 15,
                         sex = 1, marital = 1,
                         dependents = 0, era = 1)
data_fixed_fac2.0no <- data_fixed2.0no


#convert all to data_fixed_fac factors matching model
data_fixed_fac2.0no$treatedyn <- factor(data_fixed_fac2.0no$treatedyn, labels =
                                   c("No"))

data_fixed_fac2.0no$evermj <- factor(data_fixed_fac2.0no$evermj, labels =
                                  c("yes"))

data_fixed_fac2.0no$evercoc <- factor(data_fixed_fac2.0no$evercoc, labels =
                                   c("yes"))

data_fixed_fac2.0no$edu4cat <- factor(data_fixed_fac2.0no$edu4cat, labels = c("<HS", "C Grad"))

data_fixed_fac2.0no$age <-factor(data_fixed_fac2.0no$age, labels =
                              c(" 35-49"))

data_fixed_fac2.0no$sex <- factor(data_fixed_fac2.0no$sex, labels = c(" m"))

data_fixed_fac2.0no$marital <-factor(data_fixed_fac2.0no$marital, labels =
                                  c(" married"))

data_fixed_fac2.0no$dependents <-factor(data_fixed_fac2.0no$dependents, labels =
                                     c(" 0"))

#Determine prediction for fixed characteristics and treated (none, distant, recent)
prediction2.0no <- predict.glm(nb_famincyn, data_fixed_fac2.0no)

#determine probability_dnbinom for each prediction
prob_dnbinom2.0no <- 1 - (prediction2.0no/ (prediction2.0no + 1))

#run dnbinom for each probability[1:3]
hsn_dnbinom <- dnbinom(0:10, 1, prob_dnbinom2.0no[1])

cgradn_dnbinom <- dnbinom(0:10, 1, prob_dnbinom2.0no[2])

#create dataframe for graph for each none/distant/recent_dnbinom
hsn_data_dnbinom <- data.frame(dr5day = 0:10, dense = hsn_dnbinom,
                               edu4cat = "<HS", treatedyn = "No Treatment")

cgradn_data_dnbinom <- data.frame(dr5day = 0:10, dense = cgradn_dnbinom,
                                  edu4cat = "C Grad", treatedyn = "No Treatment")
```

```r
#stack dataframes into data_dnbinom

datan_dnbinom <-rbind(hsn_data_dnbinom, cgradn_data_dnbinom)



#for those who attended treatment
data_fixed2.0yes <- data.frame(faminc_mid = mean(data_clean$faminc_mid), treatedyn = 1,
                               alctry = mean(data_clean$alctry),
                               evermj = 2, evercoc = 2, edu4cat = 1:2, age = 15, sex = 1, marital = 1,
                               dependents = 0, era = 1)
data_fixed_fac2.0yes <- data_fixed2.0yes


#convert all to data_fixed_fac factors matching model
data_fixed_fac2.0yes$treatedyn <- factor(data_fixed_fac2.0yes$treatedyn, labels =
                                         c("Yes"))

data_fixed_fac2.0yes$evermj <- factor(data_fixed_fac2.0yes$evermj, labels =
                                      c("yes"))

data_fixed_fac2.0yes$evercoc <- factor(data_fixed_fac2.0yes$evercoc, labels =
                                       c("yes"))

data_fixed_fac2.0yes$edu4cat <- factor(data_fixed_fac2.0yes$edu4cat, labels = c("<HS", "C Grad"))

data_fixed_fac2.0yes$age <-factor(data_fixed_fac2.0yes$age, labels =
                                  c(" 35-49"))

data_fixed_fac2.0yes$sex <- factor(data_fixed_fac2.0yes$sex, labels = c(" m"))

data_fixed_fac2.0yes$marital <-factor(data_fixed_fac2.0yes$marital, labels =
                                      c(" married"))

data_fixed_fac2.0yes$dependents <-factor(data_fixed_fac2.0yes$dependents, labels =
                                         c(" 0"))

#Determine prediction for fixed characteristics and treated (none, distant, recent)
prediction2.0yes <- predict.glm(nb_famincyn, data_fixed_fac2.0yes)


#determine probability_dnbinom for each prediction
prob_dnbinom2.0yes <- 1- (prediction2.0yes/ (prediction2.0yes + 1))


#run dnbinom for each probability[1:3]
hsy_dnbinom <- dnbinom(0:10, 1, prob_dnbinom2.0yes[1])

cgrady_dnbinom <- dnbinom(0:10, 1, prob_dnbinom2.0yes[2])

#create dataframe for graph for each none/distant/recent_dnbinom
hsy_data_dnbinom <- data.frame(dr5day = 0:10, dense = hsy_dnbinom, edu4cat = "<HS", treatedyn = "Treatm
```

```
cgrady_data_dnbinom <- data.frame(dr5day = 0:10, dense = cgrady_dnbinom,
                                  edu4cat = "C Grad", treatedyn = "Treatment")



#stack dataframes into data_dnbinom
datay_dnbinom <-rbind(hsy_data_dnbinom, cgrady_data_dnbinom)

data2.0_dnbinom <-rbind (datan_dnbinom, datay_dnbinom)
```

```
#plot it
plot2.0_dnbinom <- ggplot(data2.0_dnbinom,
                          mapping = aes(x = dr5day, y = dense, color = edu4cat))+
  geom_smooth()+
  labs(title = "Predicted Probability of C Grad's and <HS to Binge Drink w/ and w/out Treatment", x = "
       y = "Predicted Probability")+
  facet_grid(.~treatedyn)
plot2.0_dnbinom
```

```
## `geom_smooth()` using method = 'loess'
```