Homework #6

Collin Rafferty

11/4/2021

```
load("/cloud/project/Household_Pulse_data.RData")

Household_Pulse_data$vaxx <- (Household_Pulse_data$RECVDVACC == "yes got vaxx")
is.na(Household_Pulse_data$vaxx) <- which(Household_Pulse_data$RECVDVACC == "NA")

table(Household_Pulse_data$vaxx,Household_Pulse_data$SEXUAL_ORIENTATION)
```

```
##
##            NA gay or lesbian straight bisexual something else dont know
##   FALSE   133            110     7124      272            124       174
##   TRUE    843           2228    53829     2013            737       676
```

```
table(Household_Pulse_data$vaxx,Household_Pulse_data$EEDUC)
```

```
##
##         less than hs some hs HS diploma some coll assoc deg bach deg adv deg
##   FALSE          115     269       1647      2396      1132     1565     813
##   TRUE           290     652       6097     12022      6266    18272   16727
```

```
table(Household_Pulse_data$vaxx,Household_Pulse_data$GENID_DESCRIBE)
```

```
##
##            NA  male female transgender other
##   FALSE    87  2828   4855          33   134
##   TRUE    505 23862  35218         167   574
```

```
pick_use1 <- (Household_Pulse_data$REGION == "South")
dat_use1 <- subset(Household_Pulse_data, pick_use1)

dat_use1$RECVDVACC <- droplevels(dat_use1$RECVDVACC)

model_logit1 <- glm(vaxx ~ TENURE+ GENID_DESCRIBE+ EEDUC + MS+ PUBHLTH + PRIVHLTH + SEXUAL_ORIENTATION +

require(stargazer)
```

```
## Loading required package: 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(model_logit1, type = "text", title = "Regression Output for Logit Model One")
```

```
##
```

Regression Output for Logit Model One

	Dependent variable:
	vaxx
TENUREhousing owned free and clear	0.566***
	(0.102)
TENUREhousing owned with mortgage	0.261***
	(0.095)
TENUREhousing rented	-0.013
	(0.099)
TENUREhousing occupied without rent	-0.090
	(0.189)
GENID_DESCRIBEmale	-0.890
	(0.736)
GENID_DESCRIBEfemale	-1.382*
	(0.734)
GENID_DESCRIBEtransgender	-2.210**
	(0.973)
GENID_DESCRIBEother	-1.887**
	(0.807)
EEDUCsome hs	0.248
	(0.353)
EEDUCHS diploma	0.647**
	(0.305)
EEDUCsome coll	0.940***
	(0.302)
EEDUCassoc deg	1.164***
	(0.314)
EEDUCbach deg	1.745***
	(0.311)
EEDUCadv deg	2.233***
	(0.321)
MSmarried	-0.461
	(0.793)
MSwidowed	11.635
	(160.908)

```
## MSdivorced                                                  -1.532*
##                                                             (0.869)
##
## MSseparated                                                 -2.076*
##                                                             (1.204)
##
## MSnever                                                     -1.417
##                                                             (0.878)
##
## PUBHLTHno public health ins                                 -1.051
##                                                             (0.989)
##
## PUBHLTHNA                                                   -0.588
##                                                             (0.933)
##
## PRIVHLTHno private health ins                               -0.671***
##                                                             (0.058)
##
## PRIVHLTHNA                                                  -0.365***
##                                                             (0.107)
##
## SEXUAL_ORIENTATIONgay or lesbian                            11.426
##                                                             (465.613)
##
## SEXUAL_ORIENTATIONstraight                                  -0.958
##                                                             (0.829)
##
## SEXUAL_ORIENTATIONbisexual                                  -16.410
##                                                             (882.744)
##
## SEXUAL_ORIENTATIONsomething else                            12.729
##                                                             (882.744)
##
## SEXUAL_ORIENTATIONdont know                                 0.318
##                                                             (0.536)
##
## MSmarried:SEXUAL_ORIENTATIONgay or lesbian                  -10.437
##                                                             (465.614)
##
## MSwidowed:SEXUAL_ORIENTATIONgay or lesbian                  -23.520
##                                                             (492.633)
##
## MSdivorced:SEXUAL_ORIENTATIONgay or lesbian                 -10.380
##                                                             (465.614)
##
## MSseparated:SEXUAL_ORIENTATIONgay or lesbian                -10.448
##                                                             (465.615)
##
## MSnever:SEXUAL_ORIENTATIONgay or lesbian                    -10.017
##                                                             (465.614)
##
## MSmarried:SEXUAL_ORIENTATIONstraight                        0.578
##                                                             (0.863)
##
```

```
## MSwidowed:SEXUAL_ORIENTATIONstraight                          -10.959
##                                                              (160.908)
##
## MSdivorced:SEXUAL_ORIENTATIONstraight                          1.747*
##                                                               (0.935)
##
## MSseparated:SEXUAL_ORIENTATIONstraight                         1.817
##                                                               (1.256)
##
## MSnever:SEXUAL_ORIENTATIONstraight                             1.240
##                                                               (0.943)
##
## MSmarried:SEXUAL_ORIENTATIONbisexual                          16.550
##                                                              (882.744)
##
## MSwidowed:SEXUAL_ORIENTATIONbisexual                           4.254
##                                                              (897.289)
##
## MSdivorced:SEXUAL_ORIENTATIONbisexual                         16.782
##                                                              (882.744)
##
## MSseparated:SEXUAL_ORIENTATIONbisexual                        17.627
##                                                              (882.744)
##
## MSnever:SEXUAL_ORIENTATIONbisexual                            17.059
##                                                              (882.744)
##
## MSmarried:SEXUAL_ORIENTATIONsomething else                    -13.438
##                                                              (882.744)
##
## MSwidowed:SEXUAL_ORIENTATIONsomething else                    -25.353
##                                                              (897.289)
##
## MSdivorced:SEXUAL_ORIENTATIONsomething else                   -12.086
##                                                              (882.744)
##
## MSseparated:SEXUAL_ORIENTATIONsomething else                  -10.324
##                                                              (882.745)
##
## MSnever:SEXUAL_ORIENTATIONsomething else                      -12.083
##                                                              (882.744)
##
## MSmarried:SEXUAL_ORIENTATIONdont know                          -1.149*
##                                                               (0.648)
##
## MSwidowed:SEXUAL_ORIENTATIONdont know                         -13.282
##                                                              (160.908)
##
## MSdivorced:SEXUAL_ORIENTATIONdont know                         0.201
##                                                               (0.821)
##
## MSseparated:SEXUAL_ORIENTATIONdont know                        0.026
##                                                               (1.309)
##
```

```
## MSnever:SEXUAL_ORIENTATIONdont know
##
##
## GENID_DESCRIBEmale:PUBHLTHno public health ins                  0.866
##                                                                (0.891)
##
## GENID_DESCRIBEfemale:PUBHLTHno public health ins                1.365
##                                                                (0.888)
##
## GENID_DESCRIBEtransgender:PUBHLTHno public health ins           2.990**
##                                                                (1.258)
##
## GENID_DESCRIBEother:PUBHLTHno public health ins                 1.803*
##                                                                (0.987)
##
## GENID_DESCRIBEmale:PUBHLTHNA                                    0.509
##                                                                (0.832)
##
## GENID_DESCRIBEfemale:PUBHLTHNA                                  0.771
##                                                                (0.829)
##
## GENID_DESCRIBEtransgender:PUBHLTHNA                             0.594
##                                                                (1.249)
##
## GENID_DESCRIBEother:PUBHLTHNA                                   0.761
##                                                                (0.940)
##
## EEDUCsome hs:PUBHLTHno public health ins                        -0.650
##                                                                (0.538)
##
## EEDUCHS diploma:PUBHLTHno public health ins                     -0.769*
##                                                                (0.463)
##
## EEDUCsome coll:PUBHLTHno public health ins                      -0.697
##                                                                (0.459)
##
## EEDUCassoc deg:PUBHLTHno public health ins                      -0.902*
##                                                                (0.470)
##
## EEDUCbach deg:PUBHLTHno public health ins                       -0.697
##                                                                (0.465)
##
## EEDUCadv deg:PUBHLTHno public health ins                        -0.648
##                                                                (0.475)
##
## EEDUCsome hs:PUBHLTHNA                                          -0.468
##                                                                (0.534)
##
## EEDUCHS diploma:PUBHLTHNA                                       -0.435
##                                                                (0.456)
##
## EEDUCsome coll:PUBHLTHNA                                        -0.542
##                                                                (0.451)
##
```

```
## EEDUCassoc deg:PUBHLTHNA                                          -0.849*
##                                                                    (0.468)
##
## EEDUCbach deg:PUBHLTHNA                                           -0.678
##                                                                    (0.460)
##
## EEDUCadv deg:PUBHLTHNA                                            -0.946**
##                                                                    (0.472)
##
## Constant                                                          3.009***
##                                                                    (1.068)
##
## ------------------------------------------------------------------------------
## Observations                                                        22,412
## Log Likelihood                                                   -7,767.291
## Akaike Inf. Crit.                                                15,680.580
## ==============================================================================
## Note:                                              *p<0.1; **p<0.05; ***p<0.01
```

```
require(tinytex)
```

```
## Loading required package: tinytex
```

The estimates were generally in line with what I thought they would be.The most statistically significant coefficients were also what I expected- educational attainment, private and public health insurance. However, one thing I found interesting about the estimates was how marital status affected the likelihood of someone being vaccinated. Looking at the interaction term estimate for someone who is married and identifies as gay or lesbian, it is .000029 times more likely to be vaccinated than a someone in non-same-sex couple. If you examine the estimates for never-married gay and lesbian individuals, the probability of these individuals being vaccinated is far higher than their heterosexual counterparts.

$$PV_1 = f(1 * \beta_0 + 1 * \beta_1 Sexual_{straight} + 1 * \beta_2 EEDUC_{adv-deg} + 1 * \beta_3 MS_{married} + 1 * \beta_4 GENID_{male})$$

$$PV_1 = f(3.009 - .958 + 2.333 - .461 - .890)$$

$$PV_1 = (\frac{1}{1 + e^{-}(3.009 - .958 + 2.333 - .461 - .890)}) = .954$$

$$PV_2 = f(1 * \beta_0 + 1 * \beta_1 Sexual_{gayorlesbian} + 1 * \beta_2 EEDUC_{some-hs} + 1 * \beta_3 MS_{never} + 1 * \beta_4 GENID_{female})$$

$$PV_2 = f(3.009 + 11.426 + .248 - 1.417 - 1.382)$$

$$PV_2 = (\frac{1}{1 + e^{-}(3.009 + 11.426 + .248 - 1.417 - 1.382)}) = .999$$

The logit probability that someone who lives in the south and identifies as male, straight, married, and has an advanced degree is vaccinated is 95.4%. The logit probability that someone who lives in the south and identifies as female, gay or lesbian, never married, and has only some high school is vaccinated is 99.9%. This is not what I expected. However, the largest determinant of vaccination status was sexual orientation. An individual who identifies as gay or lesbian is far more likely to be vaccinated than other sexual orientations—this factor compensated for the fact that vaccination status and educational attainment level are positively correlated.

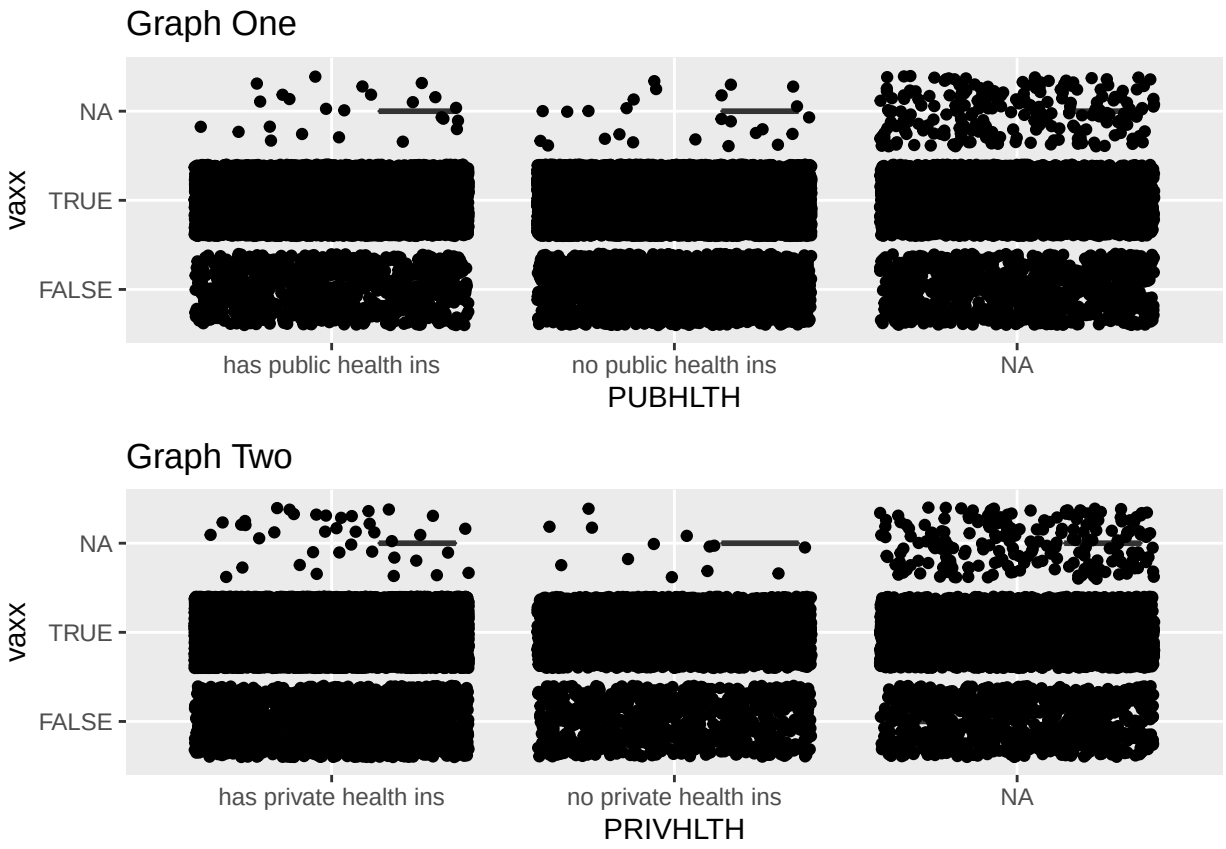
```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
require(gridExtra)
```

```
## Loading required package: gridExtra
```

```
gg1<- qplot(PUBHLTH, vaxx, data=dat_use1, geom=c("boxplot", "jitter"))+ggtitle("Graph One")
gg2<- qplot(PRIVHLTH, vaxx, data=dat_use1, geom=c("boxplot", "jitter"))+ggtitle("Graph Two")
grid.arrange(gg1,gg2)
```



I know these graphs are not that helpful, but they do provide some useful insight when examined more closely. If you look at the NA rows on both graphs, you can see the number of the NAs shown differ substantially between the two groups. It looks like there are more NAs for the group of people with private insurance. I wonder why that is?

Probit Model

```
regn_probit1 <- glm(vaxx ~ TENURE+ GENID_DESCRIBE+ EEDUC + MS+ PUBHLTH + PRIVHLTH + SEXUAL_ORIENTATION +
stargazer(regn_probit1,type="text", title = "Regression Output for Probit Model")
```

```
##
## Regression Output for Probit Model
## ================================================================================
##                                                        Dependent variable:
##                                                    ---------------------------
##                                                                vaxx
## --------------------------------------------------------------------------------
## TENUREhousing owned free and clear                            0.314***
##                                                               (0.055)
##
## TENUREhousing owned with mortgage                             0.151***
```

```
##                                                    (0.052)
##
## TENUREhousing rented                               -0.001
##                                                    (0.055)
##
## TENUREhousing occupied without rent                -0.047
##                                                    (0.107)
##
## GENID_DESCRIBEmale                                 -0.467
##                                                    (0.342)
##
## GENID_DESCRIBEfemale                              -0.705**
##                                                    (0.341)
##
## GENID_DESCRIBEtransgender                         -1.149**
##                                                    (0.508)
##
## GENID_DESCRIBEother                               -1.003**
##                                                    (0.390)
##
## EEDUCsome hs                                       0.135
##                                                    (0.210)
##
## EEDUCHS diploma                                   0.375**
##                                                    (0.182)
##
## EEDUCsome coll                                    0.531***
##                                                    (0.180)
##
## EEDUCassoc deg                                    0.655***
##                                                    (0.185)
##
## EEDUCbach deg                                     0.931***
##                                                    (0.182)
##
## EEDUCadv deg                                      1.165***
##                                                    (0.185)
##
## MSmarried                                          -0.207
##                                                    (0.381)
##
## MSwidowed                                          3.337
##                                                   (26.457)
##
## MSdivorced                                         -0.768*
##                                                    (0.435)
##
## MSseparated                                        -1.149*
##                                                    (0.678)
##
## MSnever                                            -0.759*
##                                                    (0.442)
##
## PUBHLTHno public health ins                        -0.621
```

```
##                                                          (0.501)
##
## PUBHLTHNA                                                -0.288
##                                                          (0.474)
##
## PRIVHLTHno private health ins                           -0.371***
##                                                          (0.032)
##
## PRIVHLTHNA                                              -0.192***
##                                                          (0.057)
##
## SEXUAL_ORIENTATIONgay or lesbian                          3.357
##                                                         (75.917)
##
## SEXUAL_ORIENTATIONstraight                               -0.476
##                                                          (0.405)
##
## SEXUAL_ORIENTATIONbisexual                               -5.853
##                                                         (146.955)
##
## SEXUAL_ORIENTATIONsomething else                          3.806
##                                                         (146.955)
##
## SEXUAL_ORIENTATIONdont know                               0.189
##                                                          (0.307)
##
## MSmarried:SEXUAL_ORIENTATIONgay or lesbian               -2.885
##                                                         (75.917)
##
## MSwidowed:SEXUAL_ORIENTATIONgay or lesbian               -6.915
##                                                         (80.396)
##
## MSdivorced:SEXUAL_ORIENTATIONgay or lesbian              -2.775
##                                                         (75.918)
##
## MSseparated:SEXUAL_ORIENTATIONgay or lesbian             -2.758
##                                                         (75.920)
##
## MSnever:SEXUAL_ORIENTATIONgay or lesbian                 -2.555
##                                                         (75.918)
##
## MSmarried:SEXUAL_ORIENTATIONstraight                      0.267
##                                                          (0.424)
##
## MSwidowed:SEXUAL_ORIENTATIONstraight                     -2.987
##                                                         (26.457)
##
## MSdivorced:SEXUAL_ORIENTATIONstraight                    0.880*
##                                                          (0.474)
##
## MSseparated:SEXUAL_ORIENTATIONstraight                    0.987
##                                                          (0.706)
##
## MSnever:SEXUAL_ORIENTATIONstraight                        0.661
```

```
##                                                                (0.481)
##
## MSmarried:SEXUAL_ORIENTATIONbisexual                            5.934
##                                                               (146.955)
##
## MSwidowed:SEXUAL_ORIENTATIONbisexual                            2.282
##                                                               (149.318)
##
## MSdivorced:SEXUAL_ORIENTATIONbisexual                           6.018
##                                                               (146.955)
##
## MSseparated:SEXUAL_ORIENTATIONbisexual                          6.597
##                                                               (146.956)
##
## MSnever:SEXUAL_ORIENTATIONbisexual                              6.248
##                                                               (146.955)
##
## MSmarried:SEXUAL_ORIENTATIONsomething else                     -4.240
##                                                               (146.955)
##
## MSwidowed:SEXUAL_ORIENTATIONsomething else                     -7.734
##                                                               (149.318)
##
## MSdivorced:SEXUAL_ORIENTATIONsomething else                    -3.480
##                                                               (146.955)
##
## MSseparated:SEXUAL_ORIENTATIONsomething else                   -2.383
##                                                               (146.957)
##
## MSnever:SEXUAL_ORIENTATIONsomething else                       -3.411
##                                                               (146.955)
##
## MSmarried:SEXUAL_ORIENTATIONdont know                          -0.668*
##                                                                (0.367)
##
## MSwidowed:SEXUAL_ORIENTATIONdont know                          -4.222
##                                                               (26.458)
##
## MSdivorced:SEXUAL_ORIENTATIONdont know                          0.056
##                                                                (0.467)
##
## MSseparated:SEXUAL_ORIENTATIONdont know                         0.022
##                                                                (0.790)
##
## MSnever:SEXUAL_ORIENTATIONdont know
##
##
## GENID_DESCRIBEmale:PUBHLTHno public health ins                  0.542
##                                                                (0.429)
##
## GENID_DESCRIBEfemale:PUBHLTHno public health ins               0.780*
##                                                                (0.428)
##
## GENID_DESCRIBEtransgender:PUBHLTHno public health ins          1.636**
```

```
##                                                                (0.653)
##
## GENID_DESCRIBEother:PUBHLTHno public health ins               1.035**
##                                                                (0.492)
##
## GENID_DESCRIBEmale:PUBHLTHNA                                   0.262
##                                                                (0.399)
##
## GENID_DESCRIBEfemale:PUBHLTHNA                                 0.365
##                                                                (0.397)
##
## GENID_DESCRIBEtransgender:PUBHLTHNA                            0.198
##                                                                (0.682)
##
## GENID_DESCRIBEother:PUBHLTHNA                                  0.355
##                                                                (0.473)
##
## EEDUCsome hs:PUBHLTHno public health ins                      -0.374
##                                                                (0.320)
##
## EEDUCHS diploma:PUBHLTHno public health ins                   -0.442
##                                                                (0.274)
##
## EEDUCsome coll:PUBHLTHno public health ins                    -0.381
##                                                                (0.272)
##
## EEDUCassoc deg:PUBHLTHno public health ins                    -0.493*
##                                                                (0.277)
##
## EEDUCbach deg:PUBHLTHno public health ins                     -0.346
##                                                                (0.273)
##
## EEDUCadv deg:PUBHLTHno public health ins                      -0.315
##                                                                (0.276)
##
## EEDUCsome hs:PUBHLTHNA                                        -0.269
##                                                                (0.322)
##
## EEDUCHS diploma:PUBHLTHNA                                     -0.245
##                                                                (0.274)
##
## EEDUCsome coll:PUBHLTHNA                                      -0.293
##                                                                (0.271)
##
## EEDUCassoc deg:PUBHLTHNA                                      -0.458
##                                                                (0.279)
##
## EEDUCbach deg:PUBHLTHNA                                       -0.314
##                                                                (0.274)
##
## EEDUCadv deg:PUBHLTHNA                                        -0.431
##                                                                (0.277)
##
## Constant                                                     1.615***
```

```
##                                                                                (0.512)
##
## ---------------------------------------------------------------------------------------
## Observations                                                                    22,412
## Log Likelihood                                                                -7,765.696
## Akaike Inf. Crit.                                                             15,677.390
## =======================================================================================
## Note:                                                       *p<0.1; **p<0.05; ***p<0.01
```

The signs and patterns of significance for the probit model were very similar to those of the logit model. The importance of sexual orientation as a determinant of vaccination, however, diminished.

$$PV_3 = f(1 * \beta_0 + 1 * \beta_1 Sexual_{straight} + 1 * \beta_2 EEDUC_{adv-deg} + 1 * \beta_3 MS_{married} + 1 * \beta_4 GENID_{male})$$

$$PV_4 = f(1 * \beta_0 + 1 * \beta_1 Sexual_{gayorlesbian} + 1 * \beta_2 EEDUC_{some-hs} + 1 * \beta_3 MS_{never} + 1 * \beta_4 GENID_{female})$$

```
PV_3 <- pnorm(1.615-.476+1.165-.207-.406)
PV_4 <- pnorm(1.615+3.357+.135-.759-.705)
```

The probit probability that someone who lives in the south and identifies as male, straight, married, and has an advanced degree is vaccinated is 95.4%. The logit probability that someone who lives in the south and identifies as female, gay or lesbian, never married, and has only some high school is vaccinated is 99.9%. These probabilities are identical to the logit probabilities obtained earlier. If you were not to round the probabilities to 3 digits, you would see a difference between the logit and probit estimates. Even so, they are incredibly similar to one another.

Splitting the Data into Training and Testing Sets

```
dat_use2<- dat_use1
set.seed(12345)
NN <- length(dat_use2$vaxx)
restrict_1 <- (runif(NN) < 0.3)
summary(restrict_1)
```

```
##    Mode   FALSE    TRUE
## logical   15838    6842
```

```
dat_train <- subset(dat_use2, restrict_1)
dat_test <- subset(dat_use2, !restrict_1)
pred_vals <- predict(model_logit1, type = "response")
pred_model_logit1 <- (pred_vals > 0.5)
summary(pred_model_logit1)
```

```
##    Mode   FALSE    TRUE
## logical      91   22321
```

```
pred_vals1 <- predict(regn_probit1, type = "response")
pred_regn_probit1 <- (pred_vals > 0.5)
summary(pred_regn_probit1)
```

```
##    Mode   FALSE    TRUE
## logical      91   22321
```

They appear to have the same predictive power, but I don't think that is right. So I may have to standardize some of the variables or change the dataset in some other way.