lab 6 hw 7

2022-11-04

Valery Delgado Introduction to Econometrics and Statistics Group members for the homework 7: Minghao Deng and Safinaz Ali

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

load("~/Downloads/Stats and Econometrics/Data/Household\_Pulse\_data\_w48.RData")  
attach(Household\_Pulse\_data)  
  
# Created a dataframe to analyze people’s choice to get vaxx.  
Household\_Pulse\_data$vaxx <- (Household\_Pulse\_data$RECVDVACC == "yes got vaxx")  
is.na(Household\_Pulse\_data$vaxx) <- which(Household\_Pulse\_data$RECVDVACC == "NA")   
  
sum(is.na(Household\_Pulse\_data$RECVDVACC))

## [1] 0

summary(Household\_Pulse\_data$vaxx)

## Mode FALSE TRUE NA's   
## logical 5884 40575 342

# There is a percentage of approximately 87% of people that had received the covid vaccination as opposed to approximately 13% not getting the vaccine in a total inluding the NA values of 46801.  
  
  
vaxx\_factor <- as.factor(Household\_Pulse\_data$vaxx)  
#This is used to have vaccination read as a 1 or 0 as oppossed to true or false for the logit model as vaxx is the y variable that is dependent.  
  
levels(vaxx\_factor)

## [1] "FALSE" "TRUE"

levels(vaxx\_factor) <- c("no","yes")   
# Another method to determine how R can assign levels.  
  
  
model\_logit1 <- glm(vaxx ~ EEDUC,  
 family = binomial, data = Household\_Pulse\_data)  
summary(model\_logit1)

##   
## Call:  
## glm(formula = vaxx ~ EEDUC, family = binomial, data = Household\_Pulse\_data)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.4983 0.3004 0.4099 0.6195 0.8489   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.835195 0.122852 6.798 1.06e-11 \*\*\*  
## EEDUCsome hs 0.003134 0.148254 0.021 0.9831   
## EEDUCHS diploma 0.211798 0.126724 1.671 0.0947 .   
## EEDUCsome coll 0.718288 0.125686 5.715 1.10e-08 \*\*\*  
## EEDUCassoc deg 0.699798 0.128431 5.449 5.07e-08 \*\*\*  
## EEDUCbach deg 1.599348 0.126971 12.596 < 2e-16 \*\*\*  
## EEDUCadv deg 2.240295 0.130510 17.166 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 35306 on 46458 degrees of freedom  
## Residual deviance: 32857 on 46452 degrees of freedom  
## (342 observations deleted due to missingness)  
## AIC: 32871  
##   
## Number of Fisher Scoring iterations: 5

#The model logit show that some college, associate degree, bachelors degree and advanced degree are all statistically significant in determining if someone gets a vaccine unlike a highschool education and hs diploma which are not statiscally significant.  
  
pick\_use1 <- (Household\_Pulse\_data$TBIRTH\_YEAR < 1995)   
dat\_use1 <- subset(Household\_Pulse\_data, pick\_use1)  
#These are values for R to read the vaxx factor as two levels yes or no and pick\_use1 is to look at all the variables as logic.  
  
dat\_use1$RECVDVACC <- droplevels(dat\_use1$RECVDVACC)   
  
model\_logit1 <- glm(vaxx ~ TBIRTH\_YEAR + EST\_ST + TWDAYS + INCOME + ANXIOUS+ GENID\_DESCRIBE,  
 family = binomial, data = dat\_use1)  
summary(model\_logit1)

##   
## Call:  
## glm(formula = vaxx ~ TBIRTH\_YEAR + EST\_ST + TWDAYS + INCOME +   
## ANXIOUS + GENID\_DESCRIBE, family = binomial, data = dat\_use1)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.0897 0.2626 0.3959 0.5481 1.3873   
##   
## Coefficients:  
## Estimate Std. Error  
## (Intercept) 68.198158 2.246923  
## TBIRTH\_YEAR -0.034449 0.001136  
## EST\_STAlaska 0.751738 0.170554  
## EST\_STArizona 0.346823 0.138555  
## EST\_STArkansas 0.144135 0.145218  
## EST\_STCalifornia 0.938811 0.125855  
## EST\_STColorado 0.747019 0.147065  
## EST\_STConnecticut 1.361774 0.193558  
## EST\_STDelaware 1.206422 0.222278  
## EST\_STDistrict of Columbia 2.549537 0.337746  
## EST\_STFlorida 0.660525 0.137609  
## EST\_STGeorgia 0.424317 0.139791  
## EST\_STHawaii 1.096999 0.191631  
## EST\_STIdaho 0.107615 0.143369  
## EST\_STIllinois 0.901967 0.158018  
## EST\_STIndiana 0.286494 0.143343  
## EST\_STIowa 0.407356 0.146406  
## EST\_STKansas 0.602772 0.146863  
## EST\_STKentucky 0.402486 0.151193  
## EST\_STLouisiana 0.104412 0.145099  
## EST\_STMaine 0.940435 0.213862  
## EST\_STMaryland 1.350521 0.190681  
## EST\_STMassachusetts 1.564008 0.180643  
## EST\_STMichigan 0.489201 0.137150  
## EST\_STMinnesota 0.581023 0.148127  
## EST\_STMississippi 0.218086 0.159913  
## EST\_STMissouri 0.373432 0.146411  
## EST\_STMontana -0.192642 0.157111  
## EST\_STNebraska 0.473502 0.146802  
## EST\_STNevada 0.475598 0.155835  
## EST\_STNew Hampshire 0.808706 0.166750  
## EST\_STNew Jersey 1.109311 0.179929  
## EST\_STNew Mexico 0.734368 0.157599  
## EST\_STNew York 1.321849 0.189187  
## EST\_STNorth Carolina 0.528625 0.155338  
## EST\_STNorth Dakota 0.259976 0.162243  
## EST\_STOhio 0.280215 0.145363  
## EST\_STOklahoma 0.061320 0.141157  
## EST\_STOregon 0.692383 0.142519  
## EST\_STPennsylvania 0.881924 0.154818  
## EST\_STRhode Island 1.250458 0.240148  
## EST\_STSouth Carolina 0.054645 0.143499  
## EST\_STSouth Dakota 0.327869 0.160268  
## EST\_STTennessee 0.194556 0.142496  
## EST\_STTexas 0.452985 0.123540  
## EST\_STUtah 0.614359 0.141548  
## EST\_STVermont 1.822943 0.278292  
## EST\_STVirginia 1.127621 0.160704  
## EST\_STWashington 1.176202 0.141660  
## EST\_STWest Virginia 0.404954 0.171322  
## EST\_STWisconsin 0.662359 0.156800  
## EST\_STWyoming -0.303422 0.147305  
## TWDAYShad 1-2 telework days in past week 0.778810 0.095117  
## TWDAYShad 3-4 telework days in past week 1.055850 0.108127  
## TWDAYShad 5+ telework days in past week 0.906738 0.084919  
## TWDAYShad no telework days in past week 0.234921 0.072851  
## INCOMEHH income less than $25k -0.277104 0.066275  
## INCOMEHH income $25k - $34.9k 0.057237 0.072660  
## INCOMEHH income $35k - 49.9 0.210016 0.070231  
## INCOMEHH income $50k - 74.9 0.417161 0.065267  
## INCOMEHH income $75 - 99.9 0.598192 0.070219  
## INCOMEHH income $100k - 149 0.818933 0.069918  
## INCOMEHH income $150 - 199 1.111886 0.094473  
## INCOMEHH income $200k + 1.216917 0.096254  
## ANXIOUSno anxiety over past 2 wks -0.219078 0.075920  
## ANXIOUSseveral days anxiety over past 2 wks 0.050103 0.077920  
## ANXIOUSmore than half the days anxiety over past 2 wks 0.012351 0.087640  
## ANXIOUSnearly every day anxiety -0.090630 0.083909  
## GENID\_DESCRIBEmale 0.373244 0.166574  
## GENID\_DESCRIBEfemale 0.435044 0.166010  
## GENID\_DESCRIBEtransgender 1.061478 0.342656  
## GENID\_DESCRIBEother 0.310467 0.214999  
## z value Pr(>|z|)   
## (Intercept) 30.352 < 2e-16 \*\*\*  
## TBIRTH\_YEAR -30.316 < 2e-16 \*\*\*  
## EST\_STAlaska 4.408 1.05e-05 \*\*\*  
## EST\_STArizona 2.503 0.012310 \*   
## EST\_STArkansas 0.993 0.320933   
## EST\_STCalifornia 7.459 8.69e-14 \*\*\*  
## EST\_STColorado 5.080 3.78e-07 \*\*\*  
## EST\_STConnecticut 7.035 1.99e-12 \*\*\*  
## EST\_STDelaware 5.428 5.71e-08 \*\*\*  
## EST\_STDistrict of Columbia 7.549 4.40e-14 \*\*\*  
## EST\_STFlorida 4.800 1.59e-06 \*\*\*  
## EST\_STGeorgia 3.035 0.002402 \*\*   
## EST\_STHawaii 5.725 1.04e-08 \*\*\*  
## EST\_STIdaho 0.751 0.452886   
## EST\_STIllinois 5.708 1.14e-08 \*\*\*  
## EST\_STIndiana 1.999 0.045645 \*   
## EST\_STIowa 2.782 0.005396 \*\*   
## EST\_STKansas 4.104 4.05e-05 \*\*\*  
## EST\_STKentucky 2.662 0.007766 \*\*   
## EST\_STLouisiana 0.720 0.471775   
## EST\_STMaine 4.397 1.10e-05 \*\*\*  
## EST\_STMaryland 7.083 1.41e-12 \*\*\*  
## EST\_STMassachusetts 8.658 < 2e-16 \*\*\*  
## EST\_STMichigan 3.567 0.000361 \*\*\*  
## EST\_STMinnesota 3.922 8.77e-05 \*\*\*  
## EST\_STMississippi 1.364 0.172638   
## EST\_STMissouri 2.551 0.010755 \*   
## EST\_STMontana -1.226 0.220142   
## EST\_STNebraska 3.225 0.001258 \*\*   
## EST\_STNevada 3.052 0.002274 \*\*   
## EST\_STNew Hampshire 4.850 1.24e-06 \*\*\*  
## EST\_STNew Jersey 6.165 7.04e-10 \*\*\*  
## EST\_STNew Mexico 4.660 3.17e-06 \*\*\*  
## EST\_STNew York 6.987 2.81e-12 \*\*\*  
## EST\_STNorth Carolina 3.403 0.000666 \*\*\*  
## EST\_STNorth Dakota 1.602 0.109070   
## EST\_STOhio 1.928 0.053894 .   
## EST\_STOklahoma 0.434 0.663991   
## EST\_STOregon 4.858 1.18e-06 \*\*\*  
## EST\_STPennsylvania 5.697 1.22e-08 \*\*\*  
## EST\_STRhode Island 5.207 1.92e-07 \*\*\*  
## EST\_STSouth Carolina 0.381 0.703348   
## EST\_STSouth Dakota 2.046 0.040780 \*   
## EST\_STTennessee 1.365 0.172143   
## EST\_STTexas 3.667 0.000246 \*\*\*  
## EST\_STUtah 4.340 1.42e-05 \*\*\*  
## EST\_STVermont 6.550 5.74e-11 \*\*\*  
## EST\_STVirginia 7.017 2.27e-12 \*\*\*  
## EST\_STWashington 8.303 < 2e-16 \*\*\*  
## EST\_STWest Virginia 2.364 0.018093 \*   
## EST\_STWisconsin 4.224 2.40e-05 \*\*\*  
## EST\_STWyoming -2.060 0.039415 \*   
## TWDAYShad 1-2 telework days in past week 8.188 2.66e-16 \*\*\*  
## TWDAYShad 3-4 telework days in past week 9.765 < 2e-16 \*\*\*  
## TWDAYShad 5+ telework days in past week 10.678 < 2e-16 \*\*\*  
## TWDAYShad no telework days in past week 3.225 0.001261 \*\*   
## INCOMEHH income less than $25k -4.181 2.90e-05 \*\*\*  
## INCOMEHH income $25k - $34.9k 0.788 0.430856   
## INCOMEHH income $35k - 49.9 2.990 0.002787 \*\*   
## INCOMEHH income $50k - 74.9 6.392 1.64e-10 \*\*\*  
## INCOMEHH income $75 - 99.9 8.519 < 2e-16 \*\*\*  
## INCOMEHH income $100k - 149 11.713 < 2e-16 \*\*\*  
## INCOMEHH income $150 - 199 11.769 < 2e-16 \*\*\*  
## INCOMEHH income $200k + 12.643 < 2e-16 \*\*\*  
## ANXIOUSno anxiety over past 2 wks -2.886 0.003906 \*\*   
## ANXIOUSseveral days anxiety over past 2 wks 0.643 0.520225   
## ANXIOUSmore than half the days anxiety over past 2 wks 0.141 0.887929   
## ANXIOUSnearly every day anxiety -1.080 0.280097   
## GENID\_DESCRIBEmale 2.241 0.025045 \*   
## GENID\_DESCRIBEfemale 2.621 0.008778 \*\*   
## GENID\_DESCRIBEtransgender 3.098 0.001950 \*\*   
## GENID\_DESCRIBEother 1.444 0.148728   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 32016 on 43004 degrees of freedom  
## Residual deviance: 28697 on 42933 degrees of freedom  
## (324 observations deleted due to missingness)  
## AIC: 28841  
##   
## Number of Fisher Scoring iterations: 6

new\_data\_to\_be\_predicted <- data.frame(TBIRTH\_YEAR = 1995,  
 EST\_ST = factor("New Jersey", levels = levels(dat\_use1$EST\_ST)),  
 TWDAYS = factor("had 3-4 telework days in past week",levels = levels(dat\_use1$TWDAYS)),  
 INCOME= factor("HH income $75 - 99.9",levels = levels(dat\_use1$INCOME)),  
 ANXIOUS = factor("several days anxiety over past 2 wks",levels = levels(dat\_use1$ANXIOUS)),  
 GENID\_DESCRIBE = factor("male", levels = levels(dat\_use1$GENID\_DESCRIBE))  
)  
predict(model\_logit1,new\_data\_to\_be\_predicted)

## 1   
## 2.658168

#The reason I chose the variables state, telework, income, anxious and gender is to see which one is a large contributing factor for people to get the vaccination. For instance, some states had mandatory and harsher restrictions for individuals to recieve the vaccine as New York did let go of workers that were not willing to get vaccinate.Therefore, every state had their own policies on how to handle whether the residents get vaccinated or not. Telework would be interesting to see if more days working at home vs working in person does affect a person decision in terms of getting vaccinated.Furthermore, depending on the occupation of the person some jobs have requirements and protocol on how their employees would be allowed to work or not based on their vaxx status and that can hugely affect a person income. It becomes an incentive for people to get vaccinated in order to keep their job and have a steady income especially when the pandemic closed a lot of business and had to lay off workers. When covid became known to the public there was alot of anxiety as it was a new disease and there was not any precautions, study and research to tackle the pandemic effectively. Therefore, as the first batch of doses were released alot of people got vaccinated to develop immunity against COVID 19 to lessen their anxiety as they were worried about the consquences this illness can have on their health. Laslty, in terms of gender some females were pregnant during the time of COVID 19 and were not able to get vaccinated.  
  
  
# The variables that were statiscally significant is state, telework and income as those had a variation in terms of the data. For instance California,Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, etc. other states were determined to be important for receiving vaccination based on the protocol implemented. Telework from 1-2,3-4 telework days in past week, and no telework days in past week are all statiscally significant for recieving vaccination as a person is more exposed to other. While INCOMEHH income $50k - 74.9, INCOMEHH income $75 - 99.9, INCOMEHH income $100k - 149,INCOMEHH income $150 - 199 and INCOMEHH income $200k this demonstrate as income increases a person is most likely to get vaccinated.

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.