Homework 7

safinaz alI

11/3/2022

#Literature Review #Team: Valery Delgado

Article #1:Does overcrowding and health insurance type impact patient outcomes in emergency departments? URL:<https://healtheconomicsreview.biomedcentral.com/articles/10.1186/2191-1991-3-25> Summary: As I looked over this research paper I really found it interesting because I found it to be a great sample of a research paper that talks about similar variables I will be looking at for my thesis. Also their findings were very detailed as they learned how the type of insurance you have can affect longer wait periods and raise the risk of patients experiencing bad outcomes across the board. Additionally, having insurance (private or Medicare) lessens some of the negative effects on patients’ outcomes, but not having insurance or using Medicaid worsens those effects. Additionally, the data indicate that individuals who had a main physician had a better prognosis. Furthermore, they also predicted, having an acute health condition or being older are linked to poor patient outcomes, in part because a poor result might be death. And at last, they included that having a bad outcome is more likely whether you’re black or male. Their findings are reasonable because they did run multiple restricts to test both sides of having In this research they used several methods to explain this connection by estimating a standard profit model and an instrumental variable probit model. They also ran multiple robustness checks and extended their models to test for sensitivity and a IV Probit model. These econometric techniques they did in the research paper are appropriate because they did a probit model to determine the likelihood that their variables of type of insurance with given qualities of demographic falls into a particular category of wait time in ER. Their results showed a clear statement of the impact. How time waits is either short or long, and they based it off of many factors either by demographic, health status and more which I find interesting because I would have assumed that anyone that enters the emergency room will be treated fairly since it’s sometimes life or death and based on money or the color of your skin.

Article #2:The Relationship Between Demographic, Socioeconomic, and Health-Related Parameters and the Impact of COVID-19 on 24 Regions in India: Exploratory Cross-Sectional Study URL:<https://healtheconomicsreview.biomedcentral.com/articles/10.1186/2191-1991-3-25> Summary: This research paper specifically focused their research question in India, The reason I want to use this article is because I want to see what are the different variables they used with very specific and limited data. It also gives us some background information on some international countries on how their health is doing and what it can be based on. Maybe this topic is a universal issue which is why I’m interested in their findings because it came to be a similar finding to how demographics can also affect their health issue. An exploratory cross-sectional study based on information formally published by the Indian government was employed in this study. The Ministry of Health and Family Welfare’s website was used to find out information on COVID-19.Based on the research paper the male-to-female sex ratio and the existence of an international airport in a given state were both found to have an impact on COVID-19. The sex ratio and other serious illnesses, notably heart disease, were related adversely with the death rate for COVID-19. Multiple studies showed that the sex ratio was strongly and adversely correlated with the COVID-19 crude mortality rate. Some econometric techniques they use were using both bivariate and multivariate analyses. In which they did some based stats of finding the variance to see the significant level of the cause of variables. I found this research paper accurate but not as good in running their data because they only did basic stats on the data, they did not create any regression models or any predicted probit models to help get an accurate analysis for their findings. Essentially I don’t find it appropriate as they just focus on finding a p value and seeing if it was lower or higher then the significance level of their factors. I felt as if their findings were already known or they didn’t look at enough diseases to see what other factors there can be that created a higher risk of a death rate if having covid like education status or even if vaccine was available in India that could have prevented it.

## 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("~/Desktop/Statistic & Econmetric/Data/Household\_Pulse\_data\_w48.RData")  
attach(Household\_Pulse\_data)  
  
#For this data i decided to look at factors that depict if people get vaccinated or not and some major variables was people kind of work in which you can see people who are self employed or work in a family buisiness are less likely to get vaccinated while people in private or government are likely with nonprofit company employees are more likely. another variable was region and the reason i chose that was because i felt as if when the vaccine first came out it was a limited amount in certain states however based on the regression it seems to not really play a factor as it was all negative. Lastly i looked at where people live and i realize if you lived in a one family you were less likely to get vaccinated and that could be because of the fact that you are secluded in your home versus a person in a building most likely get vaccinated because there are other people around. Lastly for race Asians were more likely to get vaccinated and that could be because of the risk factors.   
  
#creating a subgroup   
Household\_Pulse\_data$vaxx <- (Household\_Pulse\_data$RECVDVACC == "yes got vaxx")  
is.na(Household\_Pulse\_data$vaxx) <- which(Household\_Pulse\_data$RECVDVACC == "NA")   
  
 #This is my subgroup i created for age in which i focused on people over 16 since that is uaully the age to start working  
pick\_use1 <- (Household\_Pulse\_data$TBIRTH\_YEAR < 2004)   
dat\_use1 <- subset(Household\_Pulse\_data, pick\_use1)  
dat\_use1$RECVDVACC <- droplevels(dat\_use1$RECVDVACC)   
  
table(Household\_Pulse\_data$vaxx,Household\_Pulse\_data$KINDWORK)

##   
## NA work for govt work for private co work for nonprofit  
## FALSE 2449 380 2110 226  
## TRUE 16162 4244 13607 3517  
##   
## self employed work in family biz  
## FALSE 589 130  
## TRUE 2596 449

vaxx\_factor <- as.factor(Household\_Pulse\_data$vaxx)  
levels(vaxx\_factor)

## [1] "FALSE" "TRUE"

levels(vaxx\_factor) <- c("no","yes")  
  
#for the NA me and my group decided to count it as not getting vaccinated just because they did not respondits most likely because they didnt get vaccinate  
nas <- dat\_use1$RECVDVACC == 'NA'  
dat\_use1$RECVDVACC[nas] = 'no did not get vaxx'  
summary(dat\_use1$RECVDVACC)

## NA yes got vaxx no did not get vaxx   
## 0 40482 6201

#here i created a logit model to help predict the relationship between dataset variables on what im looking for specifically. glm function helped create a generalize the linear regression because the variables chosen are constrained  
model\_logit1 <- glm(vaxx ~ TBIRTH\_YEAR + KINDWORK + RRACE + REGION + LIVQTRRV, family = binomial, data = dat\_use1)  
summary(model\_logit1)

##   
## Call:  
## glm(formula = vaxx ~ TBIRTH\_YEAR + KINDWORK + RRACE + REGION +   
## LIVQTRRV, family = binomial, data = dat\_use1)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.0337 0.3143 0.4316 0.5561 1.6002   
##   
## Coefficients:  
## Estimate Std. Error z value  
## (Intercept) 62.188081 2.010094 30.938  
## TBIRTH\_YEAR -0.030559 0.001019 -29.998  
## KINDWORKwork for govt 0.805291 0.060260 13.364  
## KINDWORKwork for private co 0.269450 0.035511 7.588  
## KINDWORKwork for nonprofit 1.049835 0.074189 14.151  
## KINDWORKself employed -0.324155 0.052811 -6.138  
## KINDWORKwork in family biz -0.314629 0.107222 -2.934  
## RRACEBlack 0.044888 0.053645 0.837  
## RRACEAsian 1.562719 0.123111 12.694  
## RRACEOther -0.328926 0.056510 -5.821  
## REGIONSouth -0.710180 0.054841 -12.950  
## REGIONMidwest -0.728376 0.056690 -12.848  
## REGIONWest -0.471287 0.055813 -8.444  
## LIVQTRRVlive in mobile home -0.542670 0.066499 -8.161  
## LIVQTRRVlive in detached 1 family 0.466738 0.037074 12.589  
## LIVQTRRVlive in 1 family attached to others 0.994443 0.077573 12.819  
## LIVQTRRVlive in bldg w 2 apartments 0.642464 0.113116 5.680  
## LIVQTRRVlive in building with 3-4 apts 0.479228 0.089691 5.343  
## LIVQTRRVlive in bldg w 5+ apts 0.962041 0.059218 16.246  
## LIVQTRRVlive in boat, RV, etc -0.900528 0.143438 -6.278  
## Pr(>|z|)   
## (Intercept) < 2e-16 \*\*\*  
## TBIRTH\_YEAR < 2e-16 \*\*\*  
## KINDWORKwork for govt < 2e-16 \*\*\*  
## KINDWORKwork for private co 3.26e-14 \*\*\*  
## KINDWORKwork for nonprofit < 2e-16 \*\*\*  
## KINDWORKself employed 8.36e-10 \*\*\*  
## KINDWORKwork in family biz 0.00334 \*\*   
## RRACEBlack 0.40273   
## RRACEAsian < 2e-16 \*\*\*  
## RRACEOther 5.86e-09 \*\*\*  
## REGIONSouth < 2e-16 \*\*\*  
## REGIONMidwest < 2e-16 \*\*\*  
## REGIONWest < 2e-16 \*\*\*  
## LIVQTRRVlive in mobile home 3.33e-16 \*\*\*  
## LIVQTRRVlive in detached 1 family < 2e-16 \*\*\*  
## LIVQTRRVlive in 1 family attached to others < 2e-16 \*\*\*  
## LIVQTRRVlive in bldg w 2 apartments 1.35e-08 \*\*\*  
## LIVQTRRVlive in building with 3-4 apts 9.14e-08 \*\*\*  
## LIVQTRRVlive in bldg w 5+ apts < 2e-16 \*\*\*  
## LIVQTRRVlive in boat, RV, etc 3.43e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 35186 on 46342 degrees of freedom  
## Residual deviance: 32494 on 46323 degrees of freedom  
## (340 observations deleted due to missingness)  
## AIC: 32534  
##   
## Number of Fisher Scoring iterations: 6

#First prediction   
new\_data\_to\_be\_predicted <- data.frame(TBIRTH\_YEAR = 1990,  
 KINDWORK = factor("self employed", levels = levels(dat\_use1$KINDWORK)),  
 RRACE = factor("Black",levels = levels(dat\_use1$RRACE)),  
 REGION = factor("South",levels = levels(dat\_use1$REGION)),  
 LIVQTRRV = factor("live in bldg w 5+ apts", levels = levels(dat\_use1$LIVQTRRV))  
)  
predict(model\_logit1,new\_data\_to\_be\_predicted)

## 1   
## 1.348405

#Another prediction sample   
new\_data\_to\_be\_predicted2 <- data.frame(TBIRTH\_YEAR = 1990,  
 KINDWORK = factor("work for nonprofit", levels = levels(dat\_use1$KINDWORK)),  
 RRACE = factor("White",levels = levels(dat\_use1$RRACE)),  
 REGION = factor("Northeast",levels = levels(dat\_use1$REGION)),  
 LIVQTRRV = factor("live in detached 1 family", levels = levels(dat\_use1$LIVQTRRV))  
)  
predict(model\_logit1,new\_data\_to\_be\_predicted2)

## 1   
## 2.892383

#Based on the results prediction 2 recieved a higher outcome nased on accuracy compared to one because more of the regression had significant factors.