Homework Three [76 pts]

PA 541 - Spring 2021

Due: 4/12/2021 at 3pm

NAME:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# PART ONE

Load the data file called ‘cavax.csv’. This data contains a school-level table of rates of Personal Belief Exemptions (PBEs) in California kindergartens for the 2014-15 school year. At that time, a PBE allowed a child to enter kindergarten without having received the usual complement of vaccinations. The data contains the following variables:

**Variable name (Description)**

* *code* (School)
* *county* (County name)
* *name* (School name)
* *type* (School type (public or private))
* *district* (School district name)
* *city* (City location)
* *enrollment* (Total kindergarten enrollment)
* *pbe\_pct* (Rounded percent of personal belief exemptions)
* *exempt* (Percent of students exempt from providing vaccination records)
* *med\_exempt* (Percent exempt due to medical reasons)
* *rel\_exempt* (Percent exempt due to religious reasons)

## QUESTION 1 [8 pts]

##### a. In how many schools is the percentage of students exempt for medical reasons (med\_exempt) greater than the percentage exempt for religious reasons (rel\_exempt)? [2 pts] Of this set of schools, what percent are public schools? [2 pts]

##### b. Which county, when averaging across all the schools in that county, has the highest average percentage of exempt students (exempt)? Note, we are using the variable exempt here. [2 pts]

##### c. Create a bar chart that shows for private and public schools (type) the percent of students exempt from providing vaccination records (exempt). [2 pts]

## QUESTION 2 [10 pts]

##### Estimate a model predicting exempt (exempt) by district type (type) and enrollment (enrollment). [2 pts] Treat exempt as a continuous variable (and thus you can use standard OLS). Interpret the intercept and the coefficients. [4 pts] What is the predicted exempt percentage for a public school with 100 students in kindergarten? [2 pts] What is the predicted exempt percentage for a private school with 80 students in kindergarten? [2pts]

## QUESTION 3 [8 pts]

##### Test whether the assumption of homoskedasticy has been met. [2 pts] Discuss results. [2 pts] Calculate the VIF for each variable. [2 pts] Should we be concerned with multicollinearity. [2 pts] (Note, the vif() command is in the ‘car’ package. You can also calculate the VIF yourself as we did in class)

## QUESTION 4 [8 pts]

##### Recenter the variable enrollment at its mean. [2 pts] Create an interaction between type (type) and student enrollment (enrollment) recentered and rerun the model predicting exempt (exempt). [2 pts] Assume that type moderates the effect of enrollment in your interpretation of the interaction. (i) Interpret the results on each coefficient. (ii) Create a plot to visualize the interaction. [4 pts]

## QUESTION 5 [10 pts]

##### Let’s log transform (using the natural log) the variable enrollment and call the new variable log\_enroll. [2pts] Estimate a model predicting exempt (exempt) by district type (type) and log of enrollment (log\_enrollment). [ 2pts] Interpret the coefficient on the log of enrollment. [2 pts] Does it make more sense to use enrollment or the log of enrollment as the predictor variable? Why? [4 pts]

## QUESTION 6 [12 pts]

##### Create a binary variable to indicate high versus low exempt rates. For schools with exempt percentages equal to or greater than 33 percent, indicate them as “high”, for all other schools indicate them as “low”. [2 pts] Run a logistic regression predicting whether a school is high versus low, in other words, we want our model to predict schools falling into the high category. In your model use the predictors of school type (type) and enrollment (enrollment) (note: do **not** use log\_enroll in this model). [2 pts] Interpret the coefficients on type and enrollment in terms of both log odds and odds. [ 6 pts] What is the probability of being a high exempt school if the school is private and has 100 students enrolled? [2 pts]

# PART TWO

Read in the ‘concealed\_carry’ data. The data can be used to explore the relationship between adoption of concealed carry laws that allow to carry a concealed firearm. The data contain the following variables:

**Variable name (Description)**

* *stateid* (State id)
* *statename* (Name of state)
* *shall* (Equals 1 if state has concealed carry on that year and 0 if not)
* *year* (Year)
* *vio* (Violent crime rate per 100,000 people )
* *mur* (Murder rate per 100,000 people )

## QUESTION 7 [10 pts]

*a. Let’s begin by exploring the data. How many years are there in concealed\_carry data? How many observations per state? [2 pts]*

*b. How many states had concealed carry laws (*shall*) in 1977 and how many had concealed carry laws in 1999? [ 4pts]*

*c. Create a plot tracking the violent crime rate (*vio*) over time for states that have ever adopted conceal carry laws (*shall*) and those that have never adopted the law. [4 pts]*

## QUESTION 8 [10 pts]

##### Convert the violent crime rate (vio) into a logged variable (using the natural log), call it log\_vio. [2 pts] This will be our dependent variable. Run a pooled regression of the data (i.e., standard OLS model as if this was cross-sectionall data) predicting the log of violent crimes (log\_vio) as a function of the presence of concealed carry laws (shall) and a set of dummy variables for year. [2 pts] Iterpret the effect of shall. [2 pts] In general terms, what do the year dummy variables tell us about crime trends? [2 pts] In our current specificaiton of the model, is the effect of shall the same for all years? Why or why not? [2 pts]