Final Exam

Kyle Walter

6/21/2021

### Prep work

1. Set the working directory where the files for the final are stored and load them into the R enviroment.

setwd("R:/Graduate School/IST772 - Quantitative Reasoning for Data Science/Final Exam")  
load("allSchoolsReportStatus.RData")  
load("usVaccines.RData")  
load("districts25.RData")

1. Call the packages in that are used in the subsequent Analysis

require(data.table)

## Loading required package: data.table

require(magrittr)

## Loading required package: magrittr

require(MCMCpack)

## Loading required package: MCMCpack

## Loading required package: coda

## Loading required package: MASS

## ##  
## ## Markov Chain Monte Carlo Package (MCMCpack)

## ## Copyright (C) 2003-2021 Andrew D. Martin, Kevin M. Quinn, and Jong Hee Park

## ##  
## ## Support provided by the U.S. National Science Foundation

## ## (Grants SES-0350646 and SES-0350613)  
## ##

require(BayesFactor)

## Loading required package: BayesFactor

## Loading required package: Matrix

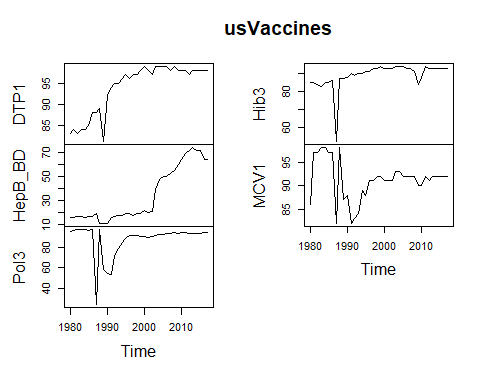
## \*\*\*\*\*\*\*\*\*\*\*\*  
## Welcome to BayesFactor 0.9.12-4.2. If you have questions, please contact Richard Morey (richarddmorey@gmail.com).  
##   
## Type BFManual() to open the manual.  
## \*\*\*\*\*\*\*\*\*\*\*\*

## Introductory/Descriptive Reports

### Question 1

1. How have U.S. vaccination rates varied over time? Are vaccination rates increasing or decreasing? Which vaccination has the highest rate at the conclusion of the time series? Which vaccination has the lowest rate at the conclusion of the time series? Which vaccine has the greatest volatility?

plot.ts(usVaccines)



Based on a quick review of the 5 vaccine that we recieved data on. It looks liek the DTP1, Hib3, and MCV1 are running the with the highest rates. DTP1 has the highest overall rate with the vaccine hovering above 95% rate. Hib3 and MCV1 are both around 90% of the US Population. We see that in all but the hepB\_BD vaccine there was some volitility in the the percentage of the populationbetween the 80s and 90s. MCV1 showing largest volitility raning from the high 95% in the mid 1980s to low 80s in th early 1990s before settling around 90% by the year 2000. While the vaccine with least amount of volitility during the the 40 period is the Hib3 which has run with 2 short dips generally between the mid 80% and low 90% of the population.It is worth noting that most of the vaccines appear to have a large, but short lived, dip in the late 80s, so this may point data quality in the information recieved.

### Question 2

1. What proportion of public schools reported vaccination data? What proportion of private schools reported vaccination data? Was there any credible difference in overall reporting proportions between public and private schools?

allSchoolsReportStatus <- as.data.table(allSchoolsReportStatus)  
allSchoolsReportStatus$nmbr <- 1  
schoolReport<- dcast(allSchoolsReportStatus, pubpriv ~ reported)

## Using 'nmbr' as value column. Use 'value.var' to override

## Aggregate function missing, defaulting to 'length'

schoolReport[,':='(Proportion = Y/sum(N, Y)), .(pubpriv)]  
schoolReport

## pubpriv N Y Proportion  
## 1: PRIVATE 252 1397 0.8471801  
## 2: PUBLIC 148 5584 0.9741800

We can see from the data that 84.7% of the private schools reported their vaccination information while 97.4% of public schools reported their vaccination rates across the US. Public schools which are funded by the states clearly show a higher propensity for reporting their vaccination numbers.

### Question 3

1. What are 2013 vaccination rates for individual vaccines (i.e., DOT, Polio, MMR, and HepB) in California public schools? How do these rates for individual vaccines in California districts compare with overall US vaccination rates (make an informal comparison to the final observations in the time series)?

districts <- as.data.table(districts)  
districts[,':='(DTP=100-WithoutDTP, Polio=100-WithoutPolio, MMR=100-WithoutMMR, HepB=100-WithoutHepB)] #inverts the without number to percentage vaccinated  
ratesDistricts <- districts[, .(avgDTP = mean(DTP), avgPolio=mean(Polio), avgMMR=mean(MMR), avgHepB=mean(HepB))]  
ratesDistricts

## avgDTP avgPolio avgMMR avgHepB  
## 1: 89.73857 90.17429 89.74 92.15714

In 2013 the average rate amoung the sample population of California Schools is as follows, 89.7% of students are vaccinated for DTP. 90.2% of students are vaccinated for Polio. 89.7% of students are vaccinated for MMR and 92.2% of students are vaccinated for HepB.

us2013Rates <- usVaccines[38,c(1,3,5,2)]#2013 Us vaccination rates  
us2013Rates

## DTP1 Pol3 MCV1 HepB\_BD   
## 98 94 92 64

Comparing the vaccines of the US to those of the sample california schools we see across the board the rates are lower in the California schools. DTP is down by a little over 8%, while Polio and measles are down by 4 and 2 percentage points respectively. The scaries one is the HepB which shows a drop of 28 percentage points.

### Question 4

1. Among districts, how are the vaccination rates for individual vaccines related? In other words, if students are missing one vaccine are they missing all of the others?

districtRates <- districts[,c(2:5)]  
cor(districtRates)

## WithoutDTP WithoutPolio WithoutMMR WithoutHepB  
## WithoutDTP 1.0000000 0.9820786 0.9779940 0.8974324  
## WithoutPolio 0.9820786 1.0000000 0.9669442 0.9127453  
## WithoutMMR 0.9779940 0.9669442 1.0000000 0.8979703  
## WithoutHepB 0.8974324 0.9127453 0.8979703 1.0000000

Based on quick correlation matrix, we can see that the likelihood of a student missing 1 vaccine than they are likely to missing all of the vaccines. DTP, Polio, and MMR have correlations near 1 while HepB is still high at near .9.

## Predictive Analyses

(For all of these analyses, use PctChildPoverty, PctFreeMeal, PctFamilyPoverty, Enrolled, and TotalSchools as predictors. Transform variables as necessary to improve prediction and/or interpretability. In general, if there is a Bayesian version of an analysis available, you are expected to run that analysis in addition to the frequentist version of the analysis.)

### Question 5

1. What variables predict whether or not a district’s reporting was complete?

First we’ll review the variables and percent complete to look for high correlation this should tell us if the variables highly follow each other

completeData <- districts[,c(7, 9:13)]  
cor(completeData)

## DistrictComplete PctChildPoverty PctFreeMeal PctFamilyPoverty  
## DistrictComplete 1.00000000 -0.04724866 -0.07442740 -0.07146700  
## PctChildPoverty -0.04724866 1.00000000 0.75422719 0.85854298  
## PctFreeMeal -0.07442740 0.75422719 1.00000000 0.72304789  
## PctFamilyPoverty -0.07146700 0.85854298 0.72304789 1.00000000  
## Enrolled -0.19074970 0.02182682 0.05963873 0.03861101  
## TotalSchools -0.21206301 0.01423951 0.05131383 0.02900816  
## Enrolled TotalSchools  
## DistrictComplete -0.19074970 -0.21206301  
## PctChildPoverty 0.02182682 0.01423951  
## PctFreeMeal 0.05963873 0.05131383  
## PctFamilyPoverty 0.03861101 0.02900816  
## Enrolled 1.00000000 0.99419685  
## TotalSchools 0.99419685 1.00000000

Based on the results we see that the variables are showing low correlation to the depent variable. There is high correlation, not surprisingly between the child poverty, free meal, and family poverty.

reportComplete<- glm(DistrictComplete~., data = completeData, family = binomial())  
summary(reportComplete)

##   
## Call:  
## glm(formula = DistrictComplete ~ ., family = binomial(), data = completeData)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.6809 0.2478 0.2911 0.3431 2.0177   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 3.6583350 0.4631718 7.898 2.82e-15 \*\*\*  
## PctChildPoverty 0.0346788 0.0319329 1.086 0.277483   
## PctFreeMeal -0.0130552 0.0110657 -1.180 0.238083   
## PctFamilyPoverty -0.0543521 0.0401800 -1.353 0.176146   
## Enrolled 0.0018799 0.0006277 2.995 0.002743 \*\*   
## TotalSchools -0.1864862 0.0548365 -3.401 0.000672 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 306.65 on 699 degrees of freedom  
## Residual deviance: 275.04 on 694 degrees of freedom  
## AIC: 287.04  
##   
## Number of Fisher Scoring iterations: 6

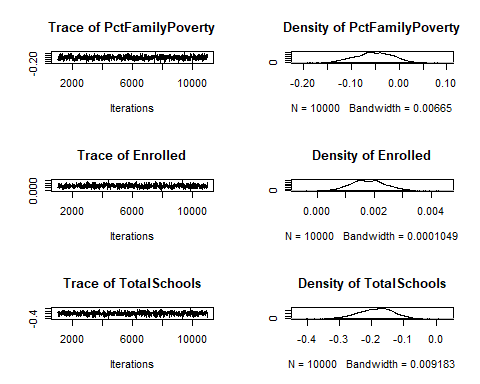
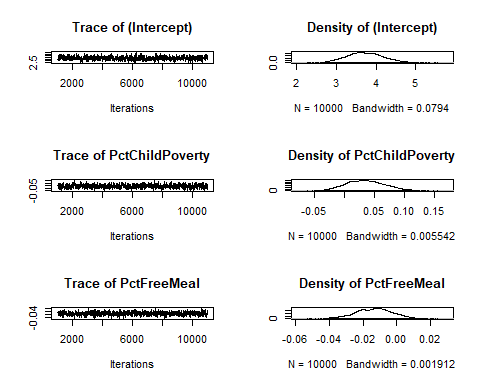
Using a frequentist mode we can see that the poverty values and percent of free meals have high P-values and fail to reject the null hypothesis that these independent variables have no effect on if the district has complete reporting. We do see the enrolled status status and total schools have p values well below the 0.05 threshold and could be used to reject the null hypothsis and show that they have impact on determining if the dependent variable actually reports completed.

Next let’s see if a bayesfactor analysis helps us determin the same thing.

BreportComplete<- MCMClogit(DistrictComplete~., data = completeData)  
summary(BreportComplete)

##   
## Iterations = 1001:11000  
## Thinning interval = 1   
## Number of chains = 1   
## Sample size per chain = 10000   
##   
## 1. Empirical mean and standard deviation for each variable,  
## plus standard error of the mean:  
##   
## Mean SD Naive SE Time-series SE  
## (Intercept) 3.725987 0.4728968 4.729e-03 2.126e-02  
## PctChildPoverty 0.036532 0.0329907 3.299e-04 1.512e-03  
## PctFreeMeal -0.013592 0.0114155 1.142e-04 5.325e-04  
## PctFamilyPoverty -0.052855 0.0395861 3.959e-04 1.745e-03  
## Enrolled 0.001809 0.0006248 6.248e-06 2.753e-05  
## TotalSchools -0.186959 0.0546590 5.466e-04 2.408e-03  
##   
## 2. Quantiles for each variable:  
##   
## 2.5% 25% 50% 75% 97.5%  
## (Intercept) 2.8416347 3.405682 3.697551 4.039025 4.691040  
## PctChildPoverty -0.0231991 0.012851 0.035080 0.058241 0.105111  
## PctFreeMeal -0.0373476 -0.021187 -0.013109 -0.005933 0.008089  
## PctFamilyPoverty -0.1295395 -0.078912 -0.052535 -0.024263 0.022085  
## Enrolled 0.0006256 0.001373 0.001781 0.002209 0.003086  
## TotalSchools -0.3015477 -0.222870 -0.183518 -0.148441 -0.089377

plot(BreportComplete)



The Bayes is reporting similar outcomes to that of the standard logistic model. The Intercept, Number Enrolled, and Density of the Total Schools in the district umbrella are the most likely predictors of the data completed. We can see that the High Density intervals do not pass through zero, but both Enrolled and Total Schools are quite close to zero. The other factors like percent of family poverty, percent of child poverty and percent of free meals served all pass through zero and do provide statistical support to show and as zero is potential value, just as in the frequentest model, we would fail to reject to the null hypothesis.

### Question 6

1. What variables predict the percentage of all enrolled students with completely up-to-date vaccines?

We’ll continue by looking at which of the indepent variables help us to predict the percentage of enrolled students who are full vacinated.

vaccinePercent <- districts[, c(6,9:13)]  
cor(vaccinePercent)

## PctUpToDate PctChildPoverty PctFreeMeal PctFamilyPoverty  
## PctUpToDate 1.0000000 0.21524588 0.28361727 0.26112174  
## PctChildPoverty 0.2152459 1.00000000 0.75422719 0.85854298  
## PctFreeMeal 0.2836173 0.75422719 1.00000000 0.72304789  
## PctFamilyPoverty 0.2611217 0.85854298 0.72304789 1.00000000  
## Enrolled 0.0646467 0.02182682 0.05963873 0.03861101  
## TotalSchools 0.0491920 0.01423951 0.05131383 0.02900816  
## Enrolled TotalSchools  
## PctUpToDate 0.06464670 0.04919200  
## PctChildPoverty 0.02182682 0.01423951  
## PctFreeMeal 0.05963873 0.05131383  
## PctFamilyPoverty 0.03861101 0.02900816  
## Enrolled 1.00000000 0.99419685  
## TotalSchools 0.99419685 1.00000000

Like before we see low correlation among the variables and percentage of students who are up to date with their vaccinations. We’ll start with frequentiest model

vaccineModel <- lm(PctUpToDate ~ ., data = vaccinePercent)  
summary(vaccineModel)

##   
## Call:  
## lm(formula = PctUpToDate ~ ., data = vaccinePercent)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -61.589 -3.385 2.931 7.132 19.382   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 81.989520 1.048703 78.182 < 2e-16 \*\*\*  
## PctChildPoverty -0.152265 0.077401 -1.967 0.049554 \*   
## PctFreeMeal 0.117317 0.028085 4.177 3.33e-05 \*\*\*  
## PctFamilyPoverty 0.315714 0.109399 2.886 0.004024 \*\*   
## Enrolled 0.006151 0.001857 3.311 0.000977 \*\*\*  
## TotalSchools -0.548971 0.171855 -3.194 0.001465 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 11.72 on 694 degrees of freedom  
## Multiple R-squared: 0.1078, Adjusted R-squared: 0.1013   
## F-statistic: 16.76 on 5 and 694 DF, p-value: 1.177e-15

The model overall looks good. We see a high F value, greater than 1 shows the the model overall works well. This is further supported by the small P-value of the of the overall model, at 1.77e-15 the value is near zero and well below the 0.05 alpha, so the model itself is such that we would reject the null hypothesis that the independent values have no effect on predicting if the percentage of students who are up to date on their vaccines.

We can also in reviewing the various variables that they all show significance in 1 way or another. Intercept, Percent Free Meal, and Enrolled are show the highest significance with values near zero. Percent Family Poverty and Total Schools also show impact with values well below the 0.05 alpha. Child Poverty is the closest variable to alpha at 0.04955 but still is less than 0.05. Thus we would reject the null hypthosis for each noting that they all have an impact on the on the dependent variable.

Checking the bayesian version of this linear regression.

BvaccineModel <- lmBF(PctUpToDate ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty + Enrolled + TotalSchools, data = vaccinePercent, posterior = T, iterations = 10000)  
summary(BvaccineModel)

##   
## Iterations = 1:10000  
## Thinning interval = 1   
## Number of chains = 1   
## Sample size per chain = 10000   
##   
## 1. Empirical mean and standard deviation for each variable,  
## plus standard error of the mean:  
##   
## Mean SD Naive SE Time-series SE  
## mu 87.818870 0.437739 4.377e-03 4.377e-03  
## PctChildPoverty -0.147524 0.076409 7.641e-04 7.641e-04  
## PctFreeMeal 0.113656 0.027797 2.780e-04 2.780e-04  
## PctFamilyPoverty 0.304592 0.108470 1.085e-03 1.085e-03  
## Enrolled 0.005903 0.001828 1.828e-05 1.875e-05  
## TotalSchools -0.526727 0.169150 1.692e-03 1.732e-03  
## sig2 137.577947 7.443666 7.444e-02 7.444e-02  
## g 0.060719 0.062239 6.224e-04 6.315e-04  
##   
## 2. Quantiles for each variable:  
##   
## 2.5% 25% 50% 75% 97.5%  
## mu 86.962918 87.519330 87.818663 88.115410 88.673753  
## PctChildPoverty -0.295866 -0.199088 -0.148106 -0.095889 0.003607  
## PctFreeMeal 0.060446 0.094619 0.113678 0.132223 0.167851  
## PctFamilyPoverty 0.091070 0.233341 0.303641 0.377005 0.518257  
## Enrolled 0.002326 0.004669 0.005908 0.007124 0.009460  
## TotalSchools -0.858908 -0.641008 -0.527255 -0.414092 -0.193806  
## sig2 123.173928 132.480764 137.350052 142.322496 152.704976  
## g 0.016327 0.030455 0.044833 0.070088 0.198868

The quantities for the variables shows good results. There is a small chance that pctFreeMeal can could pass through zero at the upper 2.5 percent but outside of it’s HDI. This is likely what caused it to show a p-value very close .05. As HDIs tend to estimate the population mean near middle using a robot, it is like the true value is in the median estimate and not in the zero, thus we can conculde we’re still able to reject the null hypthosis. In addition the other factors do not have any zero potential and thus we can also reject the null hypthosis

lmBF(PctUpToDate ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty + Enrolled + TotalSchools, data = vaccinePercent)

## Bayes factor analysis  
## --------------  
## [1] PctChildPoverty + PctFreeMeal + PctFamilyPoverty + Enrolled + TotalSchools : 2.341308e+12 ±0%  
##   
## Against denominator:  
## Intercept only   
## ---  
## Bayes factor type: BFlinearModel, JZS

We also see that the overall model has a bayes factor of 2.341308e+12. As a rule for the factor anything above a threshold of 500 has very significant support under the bayes model. Thus we would be able to reject the null hypothesis and conclude the variables do show statistically significant impact on the percentage of students who are update on their vaccinations.

### Question 7

1. What variables predict the percentage of all enrolled students with belief exceptions?

exemptions <- districts[, c(8:13)]  
cor(exemptions)

## PctBeliefExempt PctChildPoverty PctFreeMeal PctFamilyPoverty  
## PctBeliefExempt 1.00000000 -0.19582782 -0.31242141 -0.24936324  
## PctChildPoverty -0.19582782 1.00000000 0.75422719 0.85854298  
## PctFreeMeal -0.31242141 0.75422719 1.00000000 0.72304789  
## PctFamilyPoverty -0.24936324 0.85854298 0.72304789 1.00000000  
## Enrolled -0.09322614 0.02182682 0.05963873 0.03861101  
## TotalSchools -0.08263330 0.01423951 0.05131383 0.02900816  
## Enrolled TotalSchools  
## PctBeliefExempt -0.09322614 -0.08263330  
## PctChildPoverty 0.02182682 0.01423951  
## PctFreeMeal 0.05963873 0.05131383  
## PctFamilyPoverty 0.03861101 0.02900816  
## Enrolled 1.00000000 0.99419685  
## TotalSchools 0.99419685 1.00000000

For the last step of the model analysis we’ll look at students that have expressed some kind of exemption from getting vaccinated. We see that the major independent variables are have a negative correlation to the dependent variable.

next we’ll run a linear model to test prediction of the dependent variable.

exemptModel <- lm(PctBeliefExempt ~ ., data = exemptions)  
summary(exemptModel)

##   
## Call:  
## lm(formula = PctBeliefExempt ~ ., data = exemptions)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.657 -4.121 -1.901 0.735 65.519   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.349455 0.751395 13.774 < 2e-16 \*\*\*  
## PctChildPoverty 0.174444 0.055458 3.146 0.00173 \*\*   
## PctFreeMeal -0.121578 0.020123 -6.042 2.48e-09 \*\*\*  
## PctFamilyPoverty -0.219773 0.078385 -2.804 0.00519 \*\*   
## Enrolled -0.002778 0.001331 -2.088 0.03721 \*   
## TotalSchools 0.232349 0.123134 1.887 0.05958 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.399 on 694 degrees of freedom  
## Multiple R-squared: 0.1217, Adjusted R-squared: 0.1154   
## F-statistic: 19.23 on 5 and 694 DF, p-value: < 2.2e-16

In reviewing the linear model we see that all of the variables except Total Schools schools under the system rise to level of significance in predicting whether or not the students enrolled are likely to present a vaccination exemption. For Total Schools in the district we would fail to reject the null hypothesis that total schools in the district lead to a student presenting an exemption. For the other variables they present P-values that are less than 0.05 alpha, and thus we says we’re able to reject the null hypothesis.

Next we’ll run the basian version of this test and see if it tells the same story.

BexemptModel <- lmBF(PctBeliefExempt ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty + Enrolled + TotalSchools, data = exemptions, posterior = T, iterations = 10000)  
summary(BexemptModel)

##   
## Iterations = 1:10000  
## Thinning interval = 1   
## Number of chains = 1   
## Sample size per chain = 10000   
##   
## 1. Empirical mean and standard deviation for each variable,  
## plus standard error of the mean:  
##   
## Mean SD Naive SE Time-series SE  
## mu 5.731161 0.311679 3.117e-03 3.117e-03  
## PctChildPoverty 0.168454 0.054553 5.455e-04 5.455e-04  
## PctFreeMeal -0.117327 0.019961 1.996e-04 1.996e-04  
## PctFamilyPoverty -0.212921 0.076951 7.695e-04 7.564e-04  
## Enrolled -0.002693 0.001311 1.311e-05 1.311e-05  
## TotalSchools 0.225297 0.121106 1.211e-03 1.211e-03  
## sig2 70.483238 3.790205 3.790e-02 3.790e-02  
## g 0.064352 0.057254 5.725e-04 5.821e-04  
##   
## 2. Quantiles for each variable:  
##   
## 2.5% 25% 50% 75% 97.5%  
## mu 5.132409 5.517616 5.72763 5.937643 6.3468898  
## PctChildPoverty 0.060916 0.131438 0.16796 0.205872 0.2754831  
## PctFreeMeal -0.156488 -0.130692 -0.11748 -0.103985 -0.0781570  
## PctFamilyPoverty -0.363115 -0.264376 -0.21255 -0.160924 -0.0603502  
## Enrolled -0.005278 -0.003568 -0.00270 -0.001794 -0.0001413  
## TotalSchools -0.011641 0.142227 0.22599 0.305947 0.4626488  
## sig2 63.417265 67.899524 70.35887 72.980622 78.2340779  
## g 0.017407 0.032893 0.04885 0.075769 0.2066302

As we can see in the bayes model it does contain zero towards the bottom of the Total Schools model. We would similarly to the frequentest test fail to reject the null hypothesis for the total number schools as it doesn’t show significant impact on whether or not a student will present vaccine exemption. The other variables do not have HDI ranges that are passing through zero and do show significance. For these we would reject the null hypothesis and argue that they can demostrate the ability to preduct whether or not a student is likely present in exemption.

Lastly, let’s check the bayes factor of the model overall.

lmBF(PctBeliefExempt ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty + Enrolled + TotalSchools, data = exemptions)

## Bayes factor analysis  
## --------------  
## [1] PctChildPoverty + PctFreeMeal + PctFamilyPoverty + Enrolled + TotalSchools : 4.67692e+14 ±0%  
##   
## Against denominator:  
## Intercept only   
## ---  
## Bayes factor type: BFlinearModel, JZS

As we can see from the outcome of the Bayes Factor, the model show ratio of 4.67692e+14 which is significantly larger than the highest indicator on the bayes factor scale. Thus the model is significant on its own to reject the null hypothesis and say that the depend variables are showing effect on the dependent variable.

### Question 8

1. What’s the big picture, based on all of the foregoing analyses? The staff member in the state legislators’ office is interested to know how to allocate financial assistance to school districts to improve both their vaccination rates and their reporting compliance. What have you learned from the data and analyses that might inform this question?

As we can see from the first model, there is significance to show as the student population groups, schools are more likely to completely report the vaccination status of their students. However; as the number of schools in the districts grow like the likelihood of completion begins to drop. There is likely some threshold for where school must mandatory report the vaccine status of their students to the state, and as number grow schools enter that threshold. I would say the data is currently pointing to the need for better reporting resources for districts with more schools. They may need better systems to centralize their information. An alternative would be to utilize a state reporting system similar to Pennsylvania where the state can collect the information directly from the schools front line and tabulate at the state offices.

Another potential solution would be to consolidate schools, while 1 new school does cause a -.18 log drop an additional 100 students basically cancels out the effect based on the model where each additional student increases reporting likelihood by .0018 log. This can be used as the threshold for consolidation where small district are near each other.

As for improving vaccination rates, focusing on initiatives that reduce childhood poverty are the items show the most impact on increasing vaccination rates. We see in the model above that for each percent drop in childhood poverty rate within the school district we see increase in the up to date vaccination rate. We also see that as Childhood Poverty Rate falls we’re also seeing an decrease in the number of students who present with belief objects.

Family poverty rates should however; be another area to take a close look at. At the moment the model suggests that as these fall we’re getting opposite effect of vaccinations within the school population. I suspect given the recent trends for more affluent individuals to suspect a link between autism and vaccination that this has scewed the date. The people have historical been affluent and well educated and may be the reason why this trend is so strong in the model. It could also be the school districts in the same population provided include more affluent areas which leaving out the more urban dististic and can also scew the model in this way. However; I would suspect as poverty rates fall at areas and people have more disposable income to spend on things like healthcare, we’d likely see improved vaccination rates. That said just to repeat that is not what the current models are supporting.

In conclustion, focusing on childhood poverty rates and the total enrollment numbers are most likely to lead to higher vaccination and compliance in reporting based on the data provided.