IST 772 Final Exam

Andrea Brennan

Contents

[Data Overview 3](#_Toc75667319)

[usVaccines 3](#_Toc75667320)

[allSchoolsReportStatus 3](#_Toc75667321)

[Districts 3](#_Toc75667322)

[Descriptive Analysis 5](#_Toc75667323)

[Vaccination Rates Over Time 5](#_Toc75667324)

[Reporting Proportion Analysis 11](#_Toc75667325)

[California 2013 Vaccination Rates 12](#_Toc75667326)

[Correlations 13](#_Toc75667327)

[Predictive Analysis 17](#_Toc75667328)

[Reporting Rate 17](#_Toc75667329)

[Vaccine Rate 22](#_Toc75667330)

[Conclusion 27](#_Toc75667331)

[Recommendations 28](#_Toc75667332)

[Appendix A: Q&A 29](#_Toc75667333)

[Question 1a: How have U.S. vaccination rates varied over time? 29](#_Toc75667334)

[Question 1b: Are vaccination rates increasing or decreasing? 29](#_Toc75667335)

[Questions 1c: Which vaccination has the highest rate at the conclusion of the time series? 29](#_Toc75667336)

[Question 1d: Which vaccination has the lowest rate at the conclusion of the time series? 29](#_Toc75667337)

[Question 1e: Which vaccine has the greatest volatility? 30](#_Toc75667338)

[Questions 2a: What proportion of public schools reported vaccination data? 30](#_Toc75667339)

[Question 2b: What proportion of private schools reported vaccination data? 30](#_Toc75667340)

[Question 2c: Was there any credible difference in overall reporting proportions between public and private schools? 30](#_Toc75667341)

[Question 3a: What are 2013 vaccination rates for individual vaccines (i.e., DOT, Polio, MMR, and HepB) in California public schools? 30](#_Toc75667342)

[Question 3b: 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)? 30](#_Toc75667343)

[Question 4: 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? 30](#_Toc75667344)

[Question 5: What variables predict whether or not a district’s reporting was complete? 31](#_Toc75667345)

[Question 6: What variables predict the percentage of all enrolled students with completely up-to-date vaccines? 31](#_Toc75667346)

[Question 7: What variables predict the percentage of all enrolled students with belief exceptions? 31](#_Toc75667347)

[Question 8: What’s the big picture, based on all of the foregoing analyses? The staff member in the state legislator’s 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? 31](#_Toc75667348)

[Conclusion 31](#_Toc75667349)

[Recommendations 32](#_Toc75667350)

# Data Overview

Three data sets were utilized in this research. Below provides a brief summary of each.

## usVaccines

***Data Topic:***

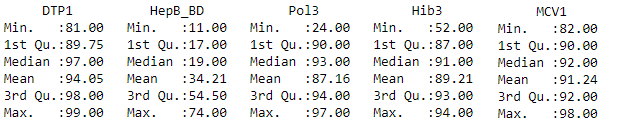
A time series data set from the World Health Organization reporting vaccination rates in the US from 1980 – 2017 for Diphtheria/Pertussis/Tetanus, Hepatitis B, Polio, Influenza, and Measles.

***Data Description:***

The data set includes 38 observations (1980 – 2017) of the following variables:

* DTP1 – vaccination rate for the first dose of the Diphtheria/Pertussis/Tetanus
* HepB\_BD – vaccination rate for hepatitis B
* Pol3 – vaccination rate for the third dose of the Polio
* Hib3 – vaccination rate for the third dose of the influenza
* MCV1 – vaccination rate for the first does of the Measles

***Data Summary:***



## allSchoolsReportStatus

***Data Topic:***

A list of California kindergartens and whether they reported vaccination data to the state in 2013.

***Data Description:***

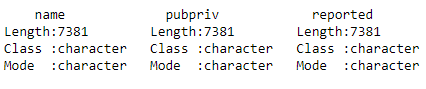
The data set includes 7,381 observations of the following 3 variables:

$ name : Name of the school

$ pubpriv : “PUBLIC” or “PRIVATE”

$ reported: “Y” or “N”

***Data Summary:***



## Districts

***Data Topic:***

A sample of California public school districts from the 2013 data collection, along with specific numbers and percentages for each district.

***Data Description:***

The data set includes 700 observations of the following 13 variables:

$ DistrictName : Name of the district

$ WithoutDTP : Percentage of students without the DTP vaccine

$ WithoutPolio : Percentage of students without the Polio vaccine

$ WithoutMMR : Percentage of students without the MMR vaccine

$ WithoutHepB : Percentage of students without the Hepatitis B vaccine

$ PctUpToDate : Percentage of all enrolled students with completely up-to-date vaccines

$ DistrictComplete: Boolean indicating whether or not the district’s reporting was complete

$ PctBeliefExempt : Percentage of all enrolled students with belief exceptions

$ PctChildPoverty : Percentage of children in the district living below the poverty line

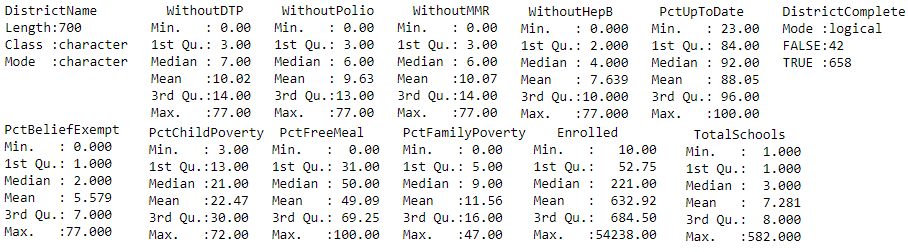
$ PctFreeMeal : Percentage of children in the district eligible for free student meals

$ PctFamilyPoverty: num Percentage of families in the district living below the poverty line

$ Enrolled : Total number of enrolled students in the district

$ TotalSchools : Total number of different schools in the district

***Data Summary:***

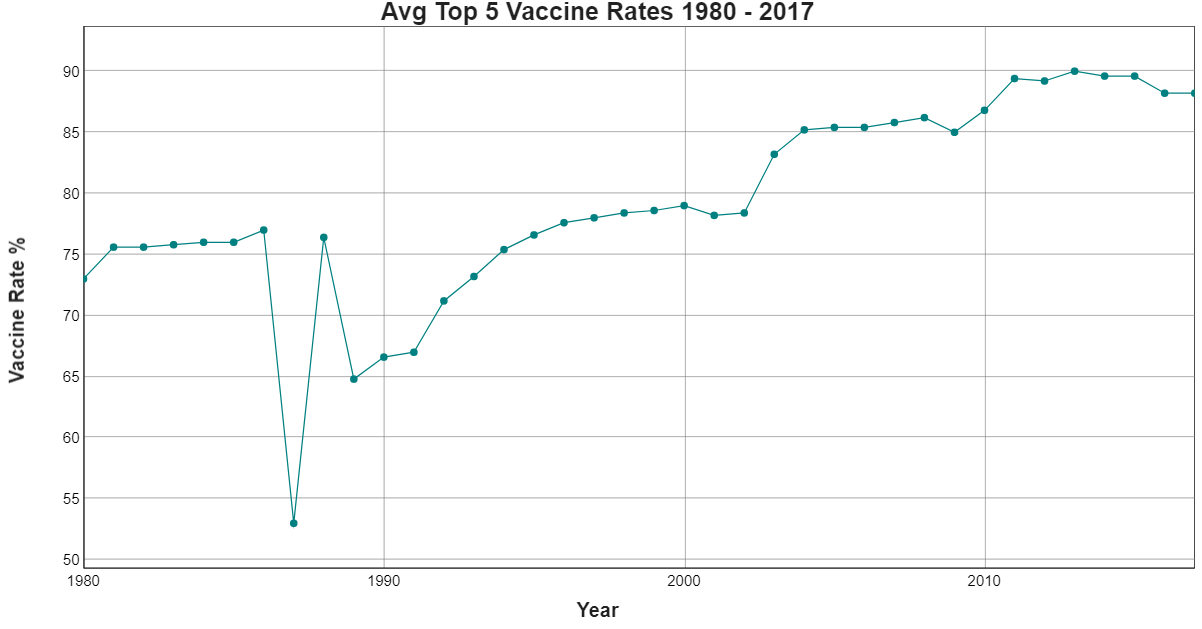


# 

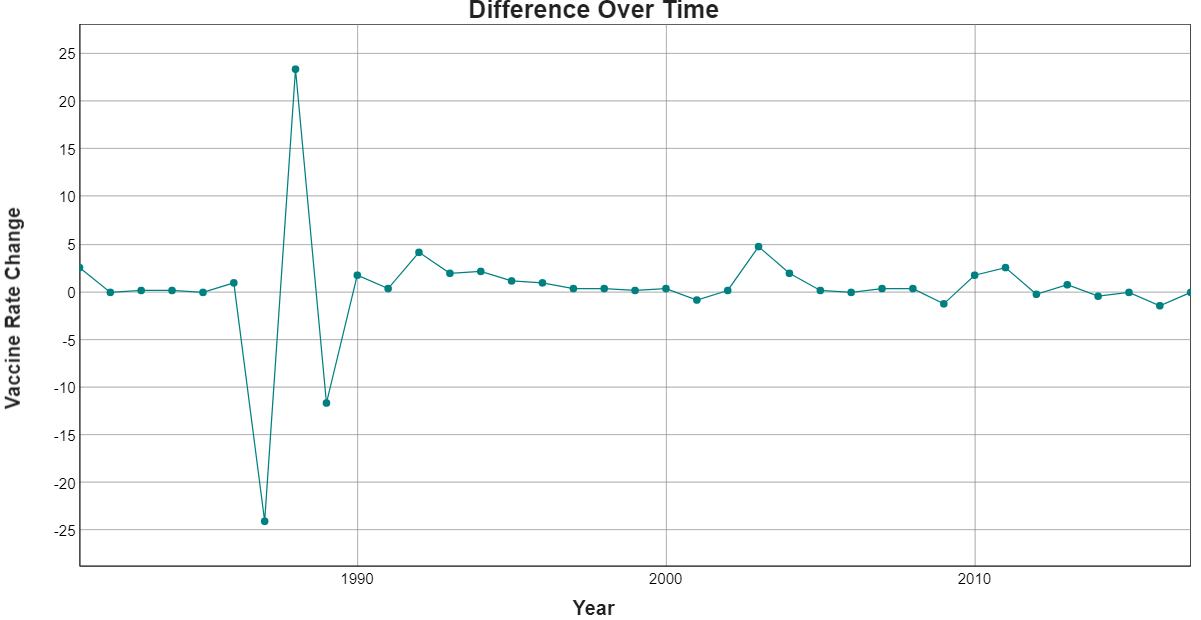
# Descriptive Analysis

## Vaccination Rates Over Time

***Average of Top 5 Vaccination Rates Combined:***



*Figure 1*

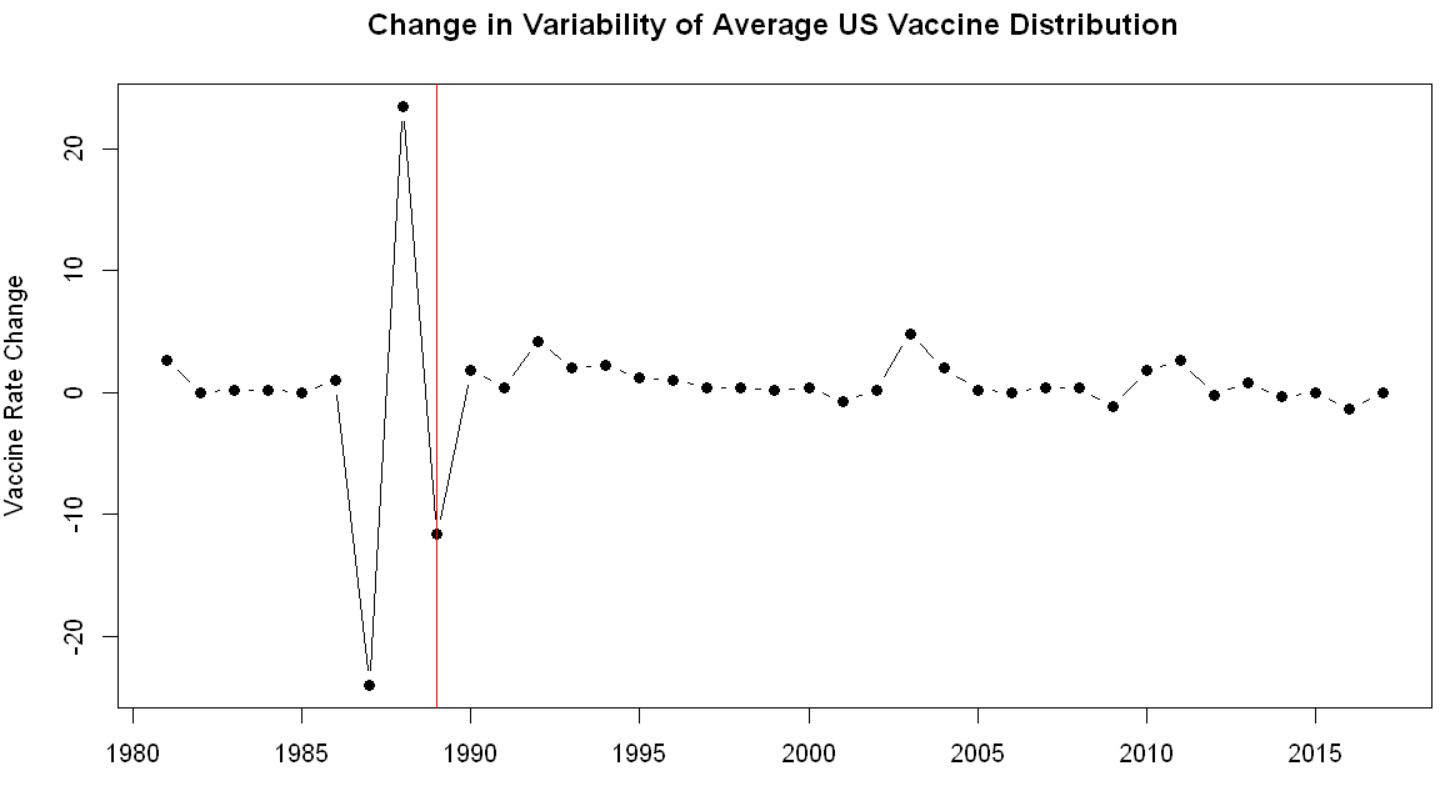


*Figure 2*

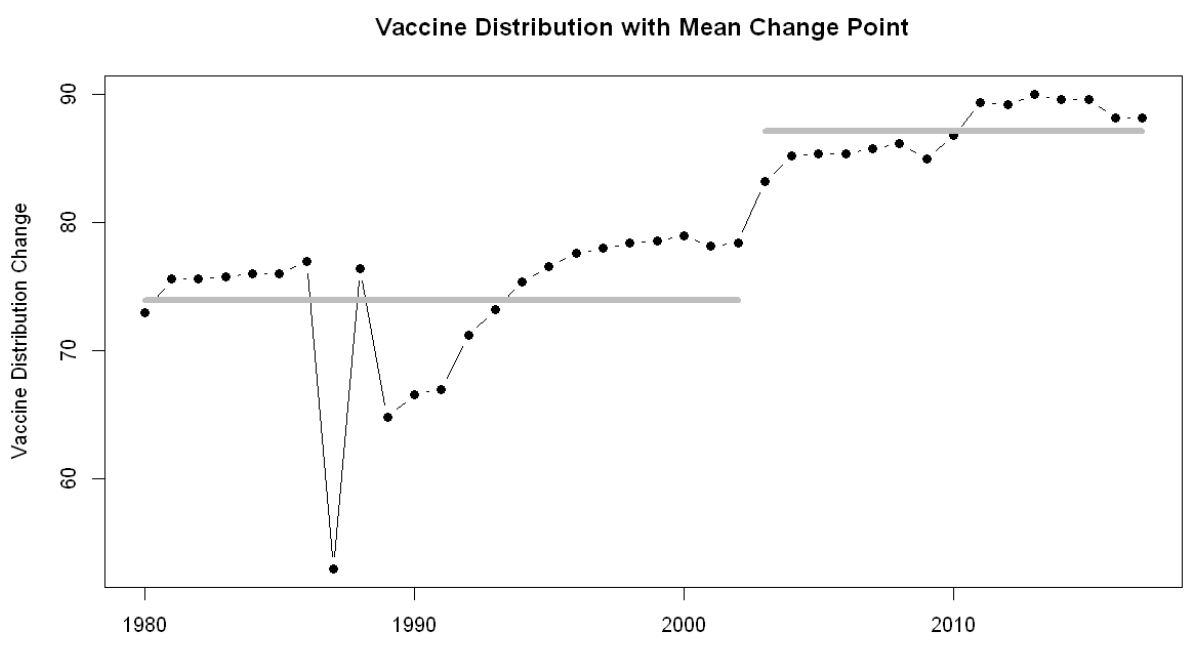
The average vaccination rates for the top 5 vaccines in the United States from 1980 to 2017 is displayed in Figure 1. In 1986 the average vaccination rate was 77%. There was an observable significant drop in the vaccination rates the following year, 1987, down to 53%, the lowest recorded for the top 5 vaccines within the data. In 1988 the rate rebounded to 75.4% before declining again the following year down to 64.8% at which point the rates resumes a less variable pattern moving forward. Overall, the vaccination rates did not recover to meet or exceed the pre-1987 rate until 1996.

Observing the differences over time in Figure 2 below removes the impacts of growth to provide a more objective view of the data. From 1980 to 1986 the range of difference for the average vaccine rate was between 0 to +2.6. In 1987 the vaccination rate did indeed plumet by -24, followed by a sharp rebound in 1988 to almost +23.4 and another drop in 1989 of -11.6. In 1990 the difference over time stabilized to be within -1.2 to +4.8.

Change point analysis of the variance further supports these findings. Figure 3 below shows the change point location of variability in 1989. See appendix for data results of change point analysis.



*Figure 3*

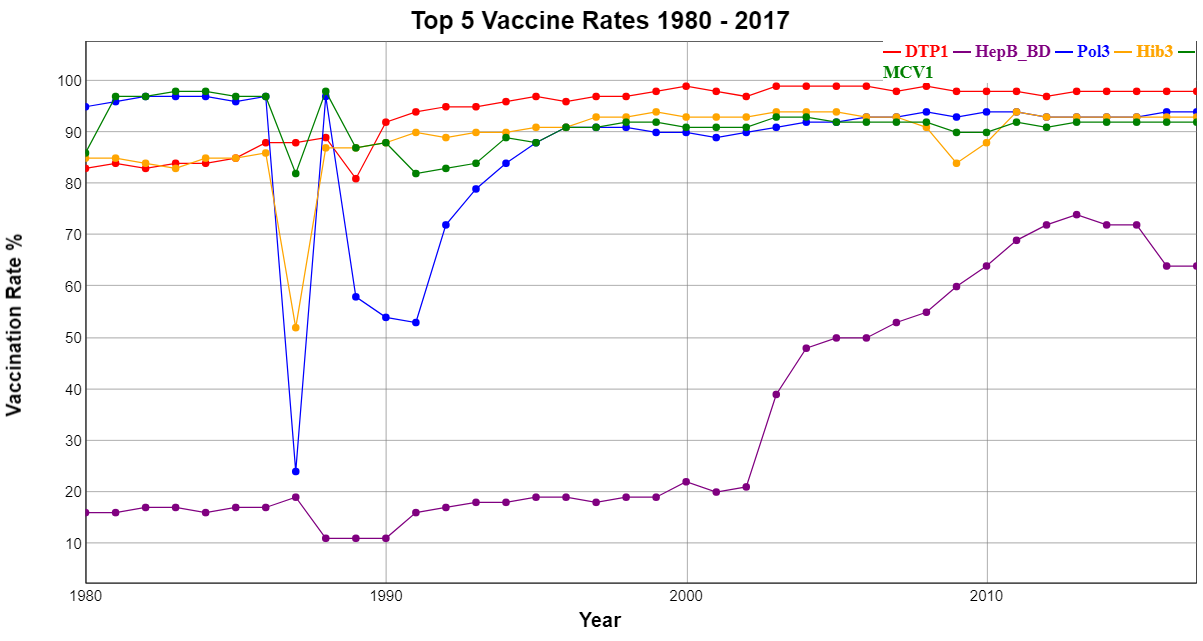


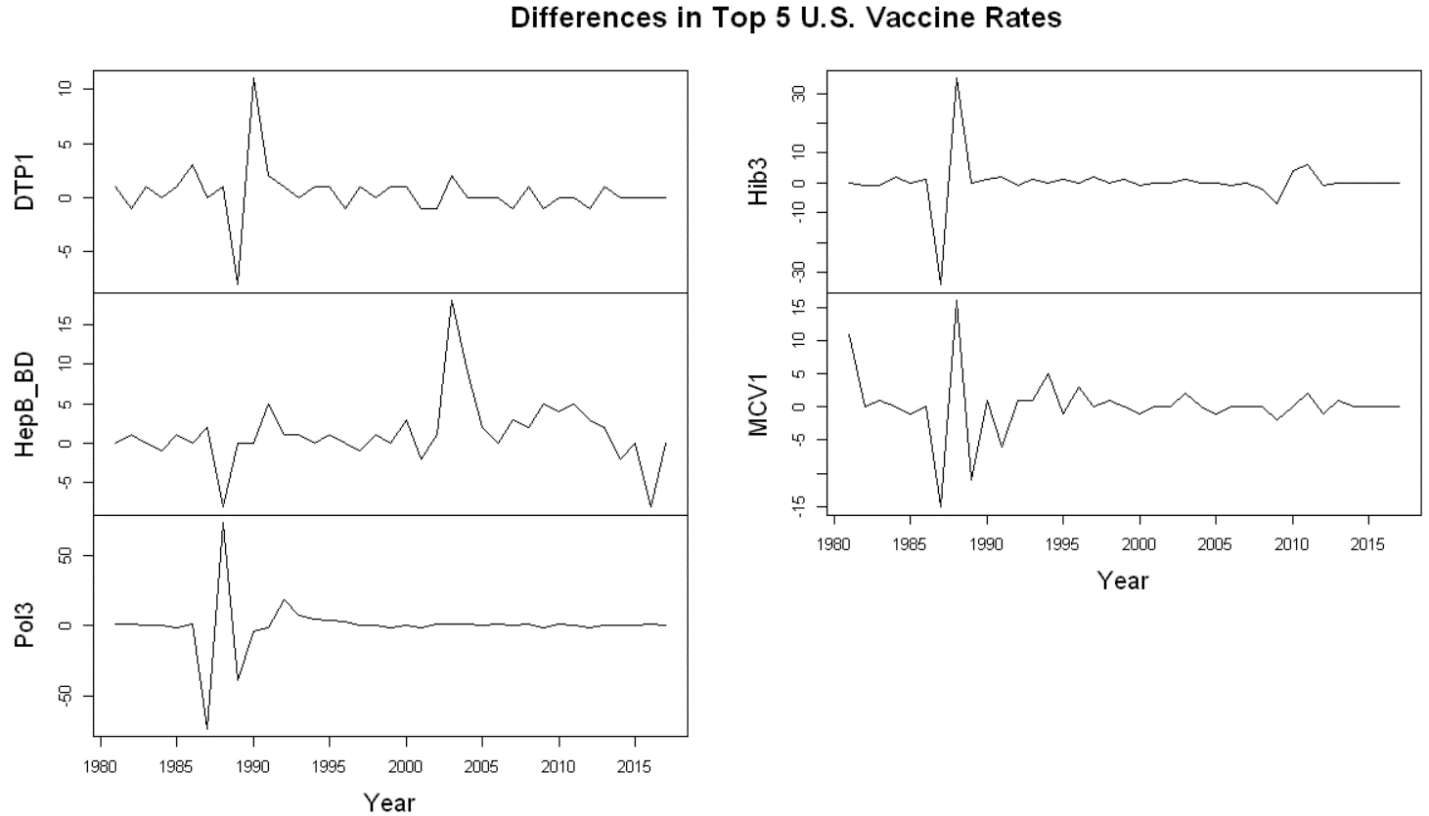
*Figure 4*

The mean change point of the average vaccine rate was location 23 as shown in Figure 4 above. This correlates to a transition from the mean of 1980 – 2002 to a new mean from 2003 – 2017. The data suggests the jump in vaccination rates in 2003 was the start to a new average range for inoculation rates. See appendix for data results of mean change point analysis. However, referring back to Figure 3 it is important to observe that since 2012 the range of difference in vaccination rates narrowed to -1.4 to + 0.8, with 9 years having less than 1% change in vaccination rates. This indicates very small growth over time.

***Top 5 Vaccination Rates:***

Breaking out the data by each of the top 5 vaccines showed additional insights into the overall trend of vaccination rates. As shown in Figure 5 below, both the Hepatitis and DTP1 vaccination rate remained stable during the 1987 vaccination rate drop while there were drops in the rates of MCV1, Hib3, and Pol3. In 1988 vaccination rates for these 3 vaccines returned to ranges close to the before the drop and then declined slightly less dramatically again in 1989. Interestingly the HepB\_BD vaccine, which had not previously been impacted by the sharp decline, suffered a decline in rates while other vaccine rates were recovering. DTP1 followed with a smaller decline in vaccination rates in 1989.

*Figure 5*



*Figure 6*

Looking at the differences in each of the top 5 vaccine rates over time in Figure 6 provided insight into the ranges of change over time that impacted the average. Individual vaccine difference ranges throughout the analyzed period are reported in Table 1 below showing the overall spread of variance for distribution of each vaccine. Note that Pol3 has the widest variability range of approximately -73 to +73 and therefore the greatest volatility.

|  |  |  |
| --- | --- | --- |
| Vaccine | Low End Of Range | Top End of range |
| DTP1 | -8 | +11 |
| HepB\_BD | -8 | +18 |
| Pol3 | -73 | +73 |
| Hib3 | -34 | +35 |
| MCV1 | -15 | +16 |

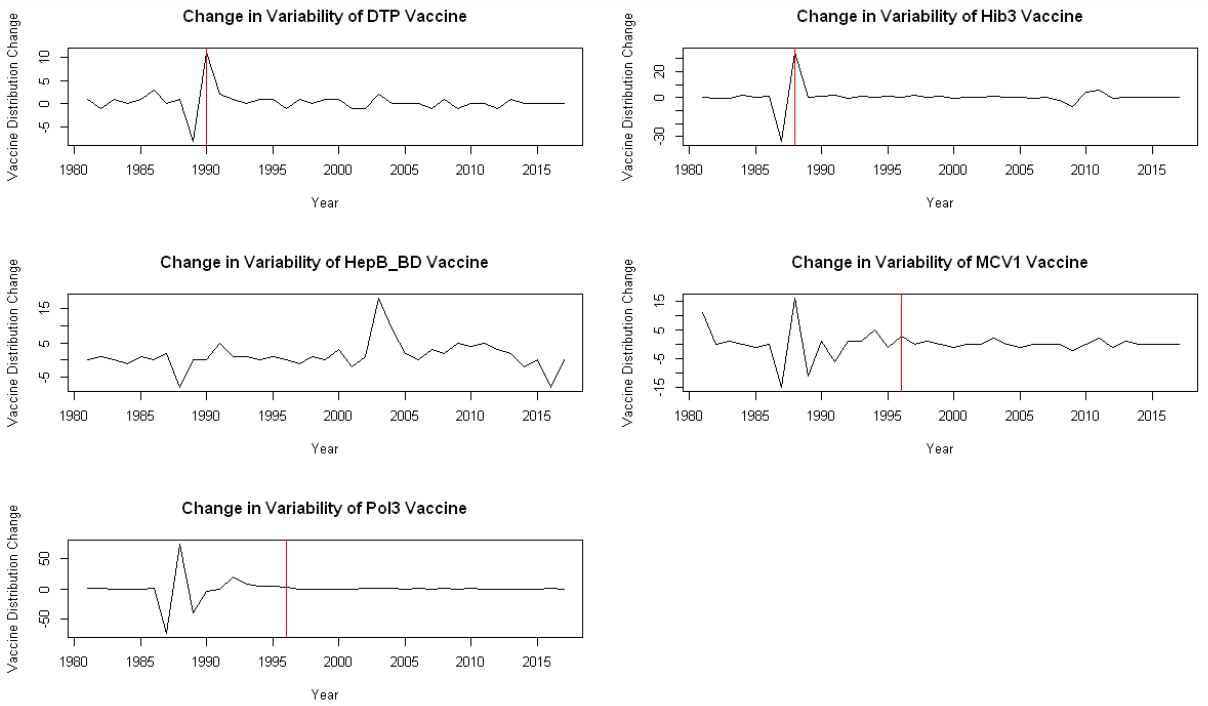
*Table 1*

##### Variability

Change in variability was observed in different points for each of the vaccines as illustrated in Table 2 and Figure 7 below. Interestingly, the HepB\_BD vaccine had no change point for variability throughout the analysis period.

|  |  |
| --- | --- |
| Vaccine | Variability change point |
| DTP1 | 1990 |
| HepB\_BD | <no observable change> |
| Pol3 | 1996 |
| Hib3 | 1988 |
| MCV1 | 1986 |

*Table 2*

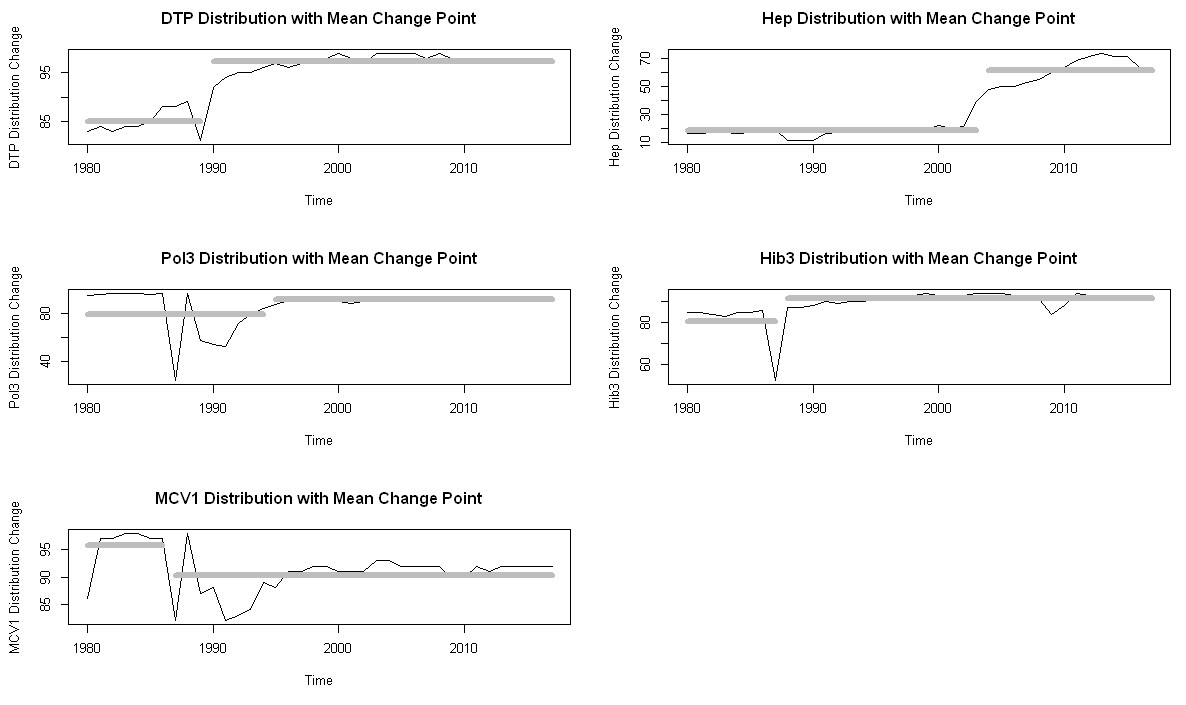


*Figure 7*

##### Change Point Analysis:

Change point analysis of the individual top 5 vaccines shows much different change points for the mean than was observed for the average. The locations of change point for the mean of DTP1, Hib3 and MCV1 were all in the late 1980s whereas the change point location for HepB was in the early 2000s and that of Pol3 was in the mid-1990s as shown below.

**Mean Change Point of Top 5 Vaccines**



*Figure 8*

**Change Point Locations of Top 5 Vaccines**

|  |  |
| --- | --- |
| Vaccine | Change point Location |
| DTP1 | 1990 |
| HepB\_BD | 2003 |
| Pol3 | 1995 |
| Hib3 | 1988 |
| MCV1 | 1986 |

*Table 3*

## Reporting Proportion Analysis

Of all the public school from the 2013 California data set, 78% were public and 22% were private. Table 4 below shows the raw data for public and private school counts reporting and not reporting vaccination data. Table 5reports the calculated percentage of the totals.

|  |  |  |  |
| --- | --- | --- | --- |
|  | PUBLIC | PRIVATE | <Total> |
| Reported Y | 5584 | 1397 | 6981 |
| Reported N | 148 | 252 | 400 |
| <Total> | 5732 | 1649 | 7381 |

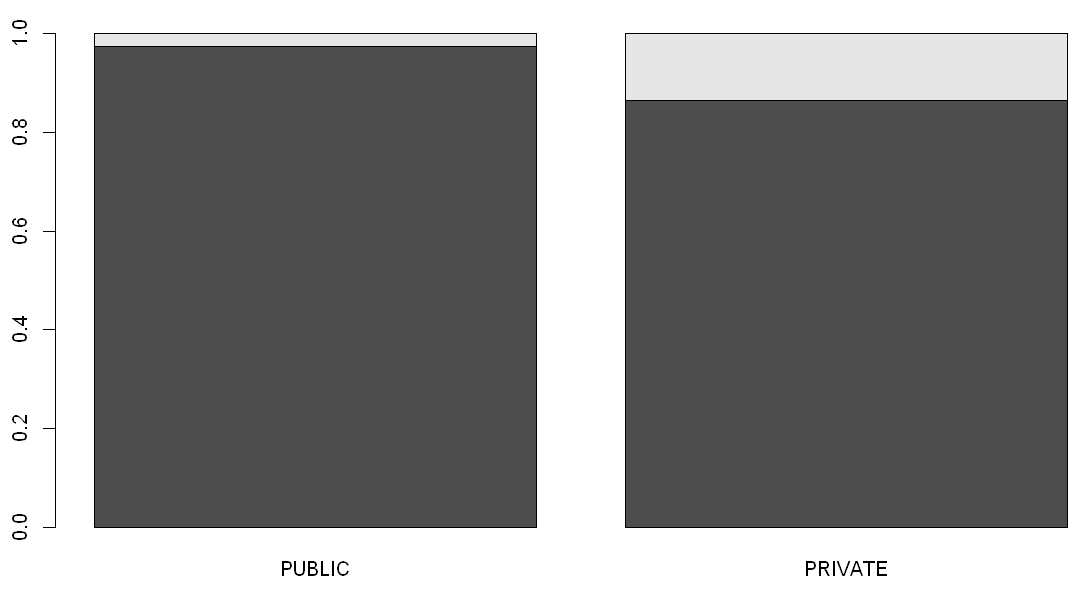
*Table 4*

Of the public schools, (0.76/0.78) 97.44% reported and (0.02/0.78) 2.56% did NOT.

|  |  |  |  |
| --- | --- | --- | --- |
|  | PUBLIC | PRIVATE | <Total> |
| Reported Y | 0.76 | 0.17 | 0.95 |
| Reported N | 0.02 | 0.03 | 0.05 |
| <Total> | 0.78 | 0.22 | 1.00 |

*Table 5*

Of the private schools, (0.19/0.22) 86.36% reported and (0.03/0.22) 13.64% did NOT.



*Figure 9*

Between public and private schools there was an 11% difference in reporting rates, with public schools reporting 97.44% and private schools reporting 86.36%.

## California 2013 Vaccination Rates

The total California public school students in 2013 was reported as 443,046. Of these, 93% had their DTP, Polio & MMR vaccines and 7% did not. 95% of children enrolled in public school in California in 2013 had their HepB vaccine and 5% did not.

**reported CALIFORNIA Students vaccinated 2013**

*(Raw Data)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DTP | Polio | MMR | HepB |
| Vaccinated | 410,821 | 412,989 | 412,009 | 421,782 |
| Not Vaccinated | 32,225 | 30,057 | 31,037 | 21,264 |

Table 6

**% of Total Reported CALIFORNIA Students Vaccinated vs National Level in 2013**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DTP | Polio | MMR | HepB |
| students Vaccinated | 93% | **93%** | **93%** | **95%** |
| students Not Vaccinated | 7% | 7% | 7% | 5% |
| NATIONAL vaccination level | **98%** | 93% | 92% | 74% |

Table 7

Nationally in 2013 ninety eight percent of Americans had received their first dose of the DPT vaccine. Ninety three percent had received the Pol3 & Hib3 vaccines, and 92% had received MCV1. HepB\_BD was the lowest distributed vaccine only being given to 74% of the US public. Interestingly, of the HepB\_BD vaccination data reported for public schools was at 95% which was higher than any of the other top 5 vaccinations reported by public schools in California for that year. This indicates that California students met or exceeded national vaccination levels for all but the DTP vaccine, having a 5% lower vaccination rate than the rest of the United States.

## Correlations

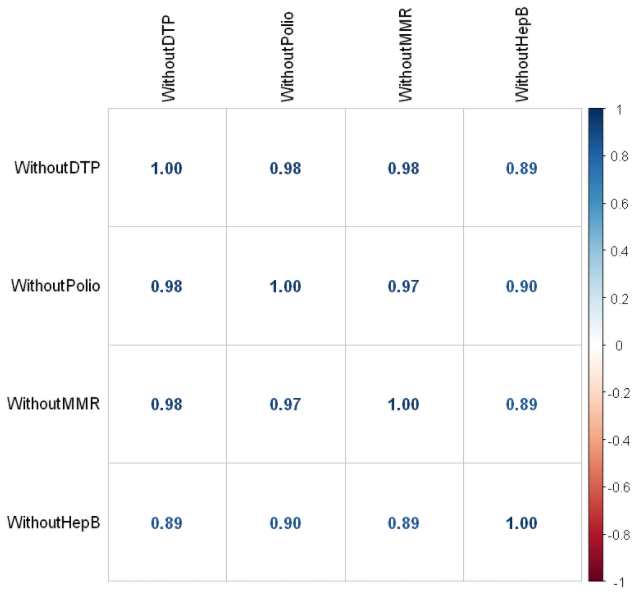
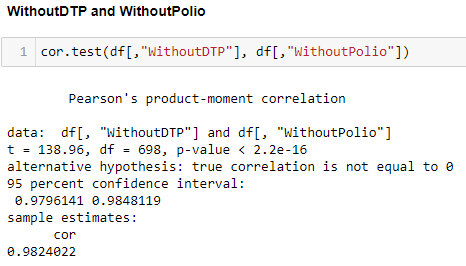


Figure 10

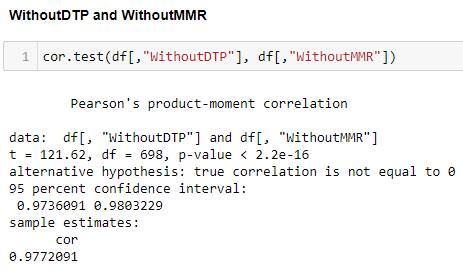
Initial correlation analysis (above) shows high positive correlation between all 4 of the variables WithoutDTP, WithoutPolio, WithoutMMR, and WithoutHepB. To further illustrate these correlations, additional Frequentist PPMC and Bayesian Correlation tests were executed between WithoutDTP, WithoutPolio, WithoutMMR & WithoutHepB – each of which represents the number of reported students who did not receive the respective vaccines. The strength of these correlations indicates there is a high correlation between students who are missing one vaccine and missing other vaccines.

#### Frequentist Analysis (PPMC)

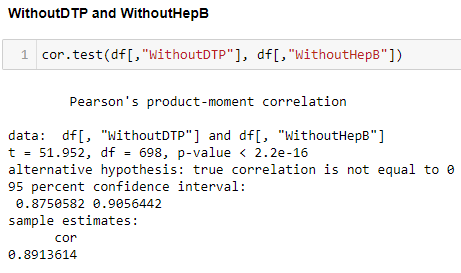
|  |  |  |  |
| --- | --- | --- | --- |
|  | WithoutPolio | WithoutMMR | WithoutHepB |
| WithoutDTP | 0.9824 | 0.9772 | 0.8914 |



The cor.test results show a t-value (138.96) with an absolute value much greater than 2 and a p-value of < 2.2e-16. Both of these indicators support a rejection of the null hypothesis that there is no correlation between the number of students who have not gotten the DTP vaccine and those who have not gotten the Polio vaccine. As the t-value is positive, it suggests a positive correlation between the two. The 95% confidence interval is 0.9796 to 0.9848 which is extremely narrow, indicating a high level of certainty. Being an entirely positive range, this provides further support for the total positive correlation of 0.9824.



The cor.test results between WithoutDTP and WithoutMMR show a t-value (121.62) with an absolute value much greater than 2 and a p-value of <2.2e-16. Both of these indicators support a rejection of the null hypothesis that there is no correlation between the number of students who have not gotten the DTP vaccine and those who have not gotten the MMR vaccine. Again, the positive t-test indicates a positive correlation and the 95% range is entirely positive between 0.9736 to 0.9803. The total correlation score is 0.9772.

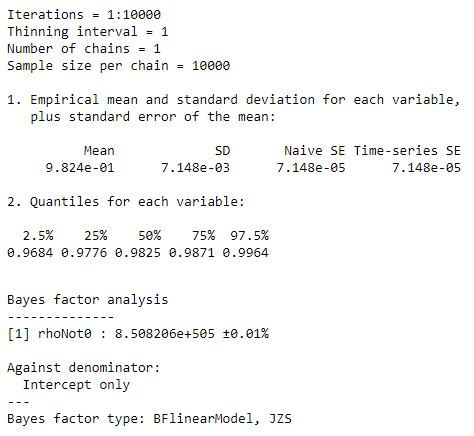


The cor.test results between WithoutDTP and WithoutHepB show a t-value (51.95) with an absolute value greater than 2 and a p-value of <2.2e-16. Both of these indicators support a rejection of the null hypothesis that there is no correlation between the number of students who have not gotten the DTP vaccine and those who have not gotten the Hepatitis B vaccine. Again, the positive t-test indicates a positive correlation and the 95% range if entirely positive between 0.8751 to 0.9056. It is worth noting the t-test value is smaller than the previous comparisons which is also reflected in the total correlation score of 0.8914.

The strength of all 3 frequentist correlations between the number of students missing one vaccine versus those missing another indicated that students who are missing one vaccine had a high correlation with not having had another vaccine.

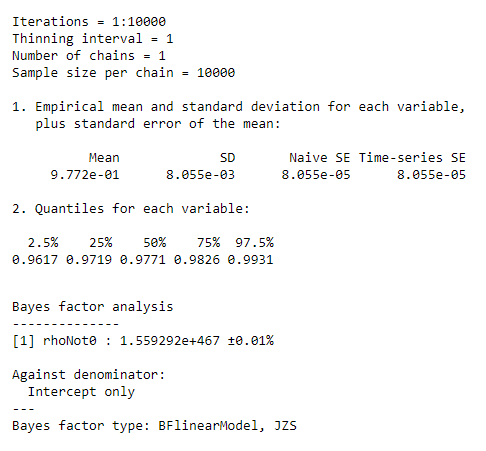
#### Bayesian Analysis

Correlation between WithoutDTP & WithoutPolio



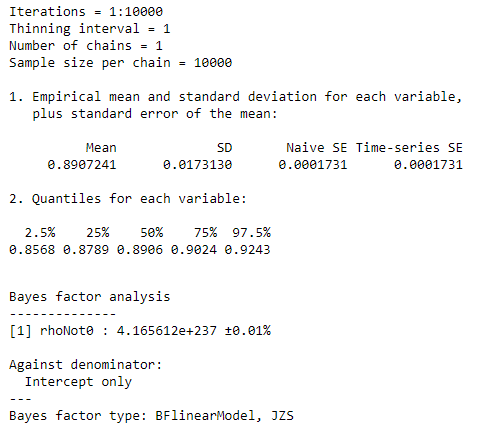
The Bayesian analysis above shows exceptionally large odds (8.508206e+505 to 1) with 0.01 margin of error that there is a correlation between not having the DTP vaccine and not having the Polio vaccine. The quantiles show a narrow 95% HDI between 0.9684 and 0.9964 supporting a significant positive correlation. This supports rejection of the null hypothesis that there is no correlation between the variables.

Correlation between WithoutDTP & Without MMR



The Bayesian correlation analysis above shows exceptionally large odds (1.559292e+446 to 1) with 0.01 margin of error that there is a correlation between not having the DTP vaccine and not having the MMR vaccine. The quantiles show a narrow 95% HDI between 0.9617 and 0.9931 supporting a significant positive correlation. This supports rejection of the null hypothesis that there is no correlation between the variables.

Correlation between WithoutDTP & WithoutHepB



The Bayesian correlation analysis above shows exceptionally large odds (4.165612e+237 to 1) with 0.01 margin of error that there is a correlation between not having the DTP vaccine and not having the HepB vaccine. The quantiles show a larger, but still narrow and entirely positive 95% HDI between 0.8568 and 0.9243 supporting a significant positive correlation. This supports rejection of the null hypothesis that there is no correlation between the variables.

The frequentist and Bayesian analysis both support the idea that students missing one vaccine in the data are far more likely to be missing another vaccine.

## Predictive Analysis

### Reporting Rate

After converting all percent variables to student counts and reviewing the distribution of districts reporting as complete (TRUE) versus incomplete (FALSE) by variable, the boxplots revealed outliers in all variables.

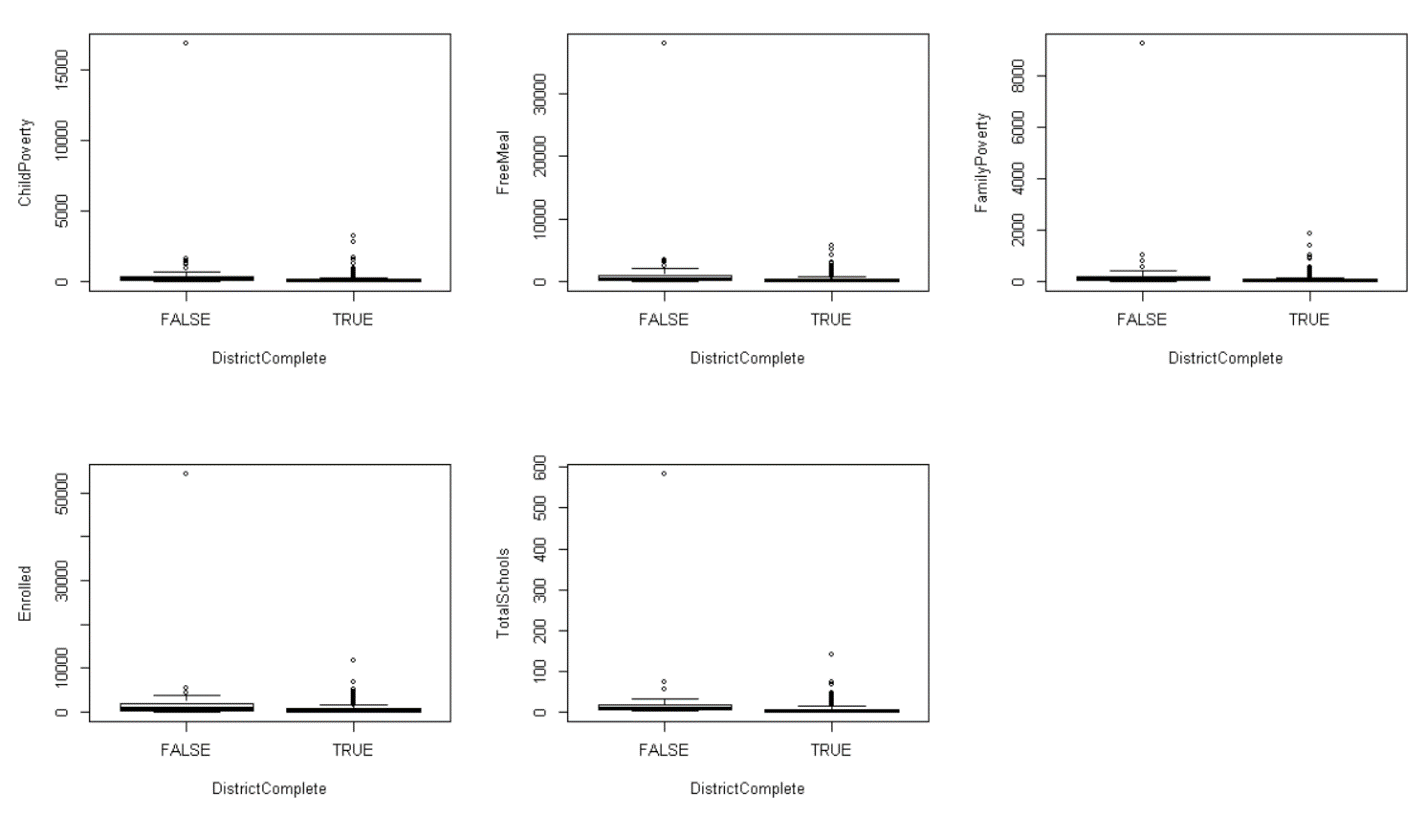
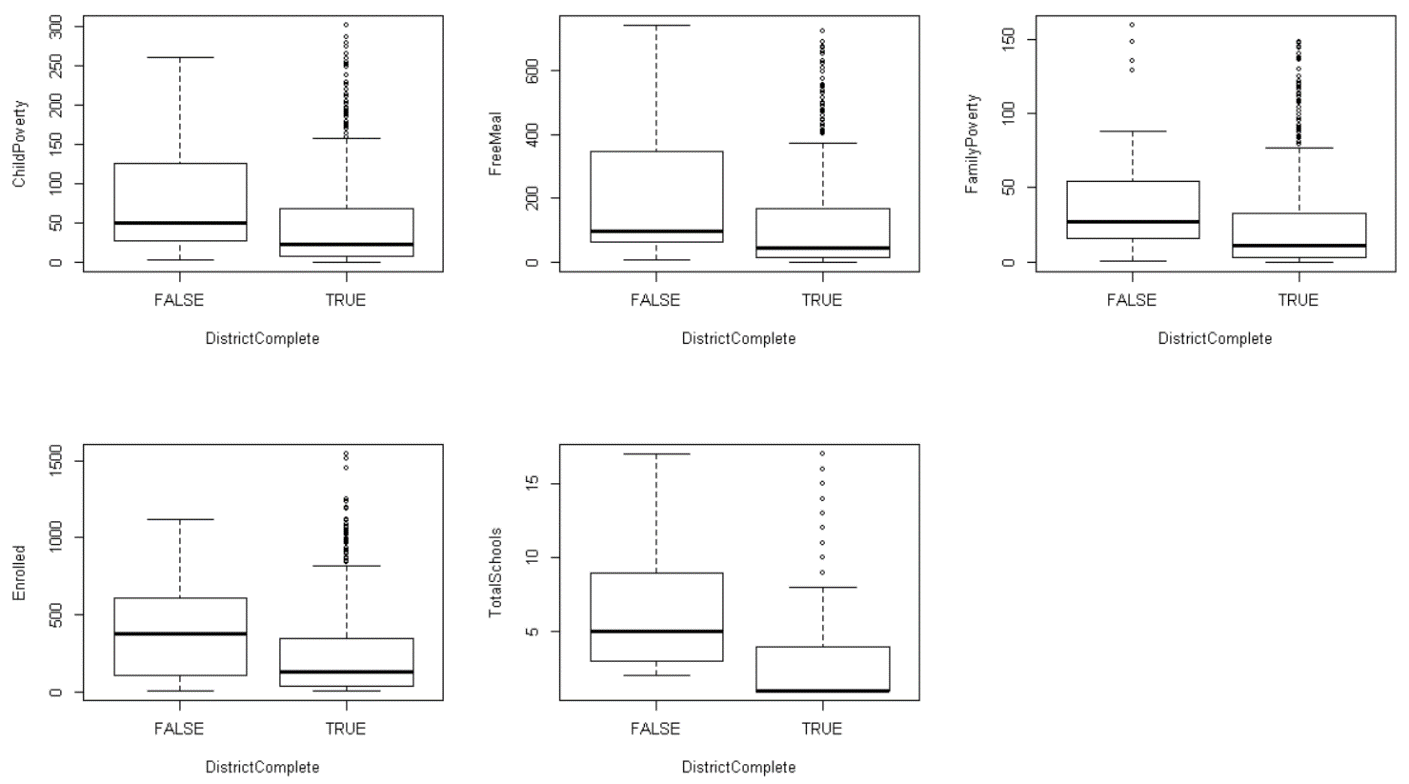


Figure 11

To reduce the potential for impacts to additional analysis, all outliers identified as anything more than 1.5\*IQR more than the 3rd quartile or 1.5\*IQR less than the first quartile for any variable were removed from the data set. Analysis of the cleaned data sets showed the distributions illustrated in Figure 12 below.

*Variable distribution after outlier removal*

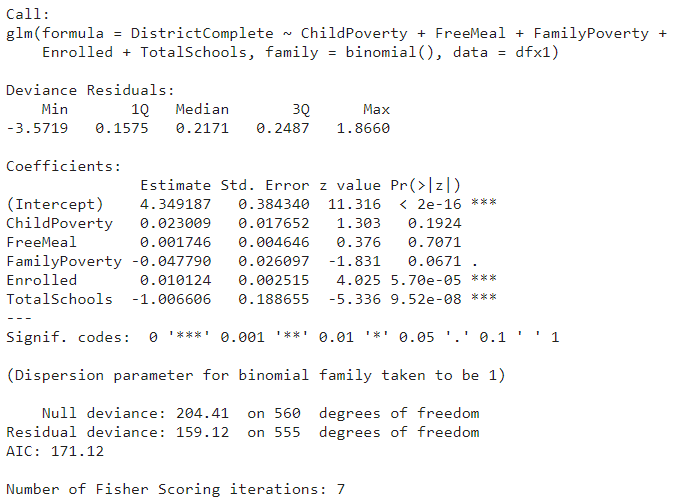


*Figure 12*

##### Logistic Regression

Logistic regression was performed to determine which variables were the strongest predictors for a district reporting completed.

*Figure 13: District Complete Logic Regression Output*



Initial logit Regression utilizing the factors of count of ChildPoverty students, count of FreeMeal students, count of FamilyPoverty students, count of students Enrolled, and TotalSchools to predict DistrictComplete resulted in an **AIC of 171.12**, showing Enrolled and TotalSchools significant at the 0.001 alpha level. The z-value of Enrolled at 4.025 was significant with the p-value of 5.70e-05, however the coefficient was close to zero at 0.01. The TotalSchools absolute z-value was significant at -5.336 along with a p-value of 9.52e-08 and a coefficient of -1.01 indicating that as the number of Total Schools decrease, the District Complete score increases.

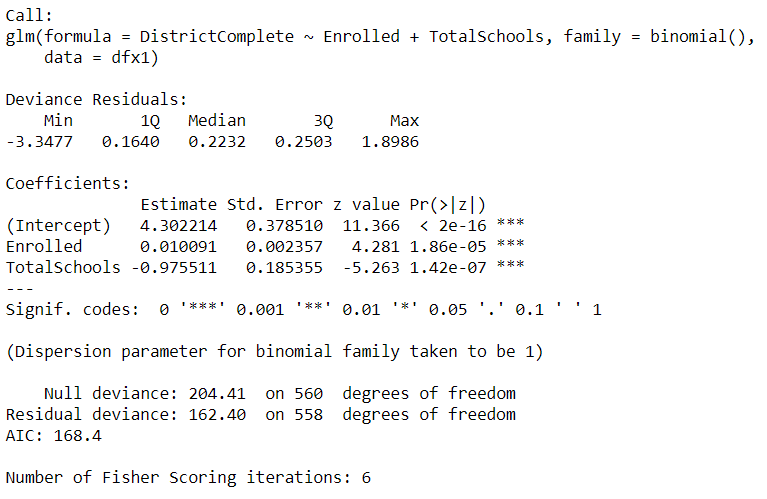
However, to be certain the best variables were selected in combination for the logistic model, another logistic regression was run independently for variable as an independent predictor of DistrictComplete (see appendix for data results). Table 8 shows the AIC score and z-score of the variable used in each independent model along with the AIC score.

|  |  |  |
| --- | --- | --- |
| Variable(s) used to predict District Outcome | Variable P-value | AIC score |
| ChildPoverty | 0.0117 | 202.92 |
| FreeMeal | 0.00876 | 202.49 |
| FamilyPoverty | 0.00625 | 201.97 |
| Enrolled | 0.00954 | 202.58 |
| TotalSchools | 6.59E-06 | 190.73 |

Table 8

After analyzing the above data, attempts at models using combinations of FamilyPoverty, Enrolled and TotalSchools were executed until the final model of DistrictComplete predicted by **enrolled students** and **total school count** was identified as the most efficient, resulting in an AIC score of **168.4.** Combining this scoring along with the z-value and p-value analysis of the combined and final model showed very strong support that the number of total schools in a district was the most important predictor of a district’s reporting, followed by the number of enrolled students. See appendix for all attempted logit regression results.

*Final model logit regression results:*



##### Chi-Square Test

A chi-square test was executed to further evaluate the strength of the chosen model using count of enrolled students and total schools as predictors of district completion of reporting.

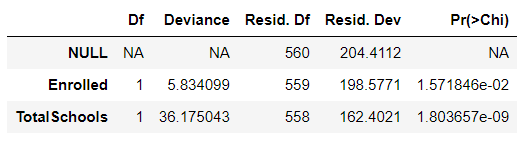


Figure 14

The p-value results below 0.01 make both chi-squared tests statistically significant. This further supports the idea that both variables are good predictors for district completion. However, it is important to note that the larger deviance of Total Schools (36.18) continues to support the idea that it is the stronger predictor of the two. Calculating the pseudo r-Squared and using the Nagelkerke results, 23.6% of the variance in the district complete variable is accounted for by the enrolled and total school variables. See appendix for data results.

To verify the fit of the **DistrictComplete ~ Enrolled + TotalSchools** model a confusion matrix was generated showing **95.26%** of predictions were accurate when using this model.

**Confusion Matrix**

|  |  |  |
| --- | --- | --- |
|  | False | True |
| Predicted False | 2 | 2 |
| Predicted True | 23 | 534 |

Table 9

##### Bayesian Estimation of Logistic Regression

Finally, to further investigate the best predictors, Bayesian estimation of logistic regression was performed on the selected model. Reviewing the log odds output (Figure 15) showed no noticeable pattern of rise or drop over time and no end being far more variable than the other end, so the initial model run with 10,000 iterations was sufficient.

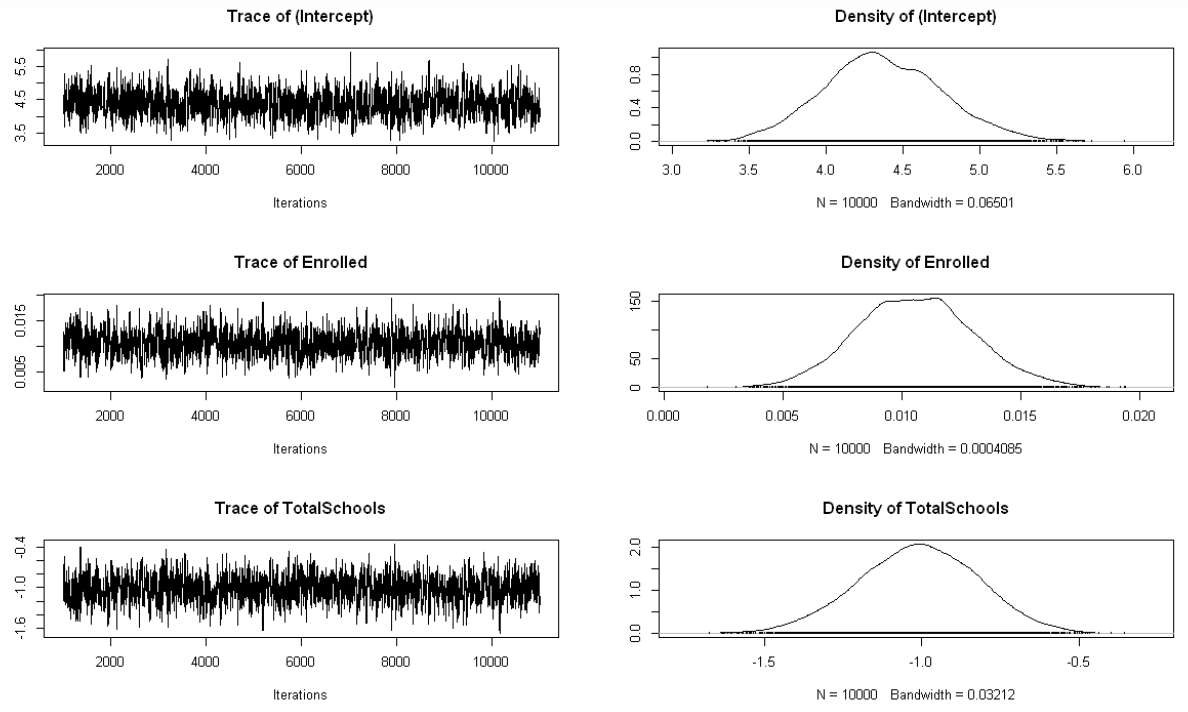
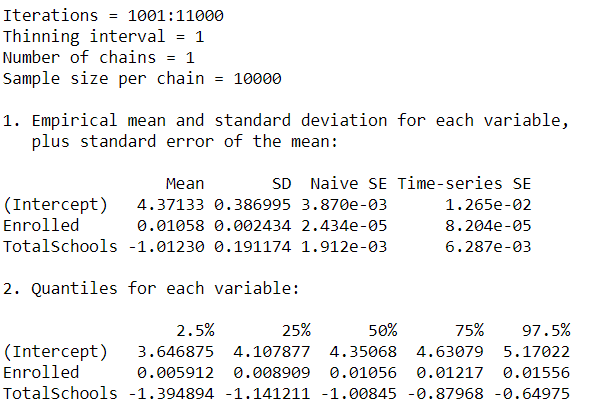
****

Figure 15

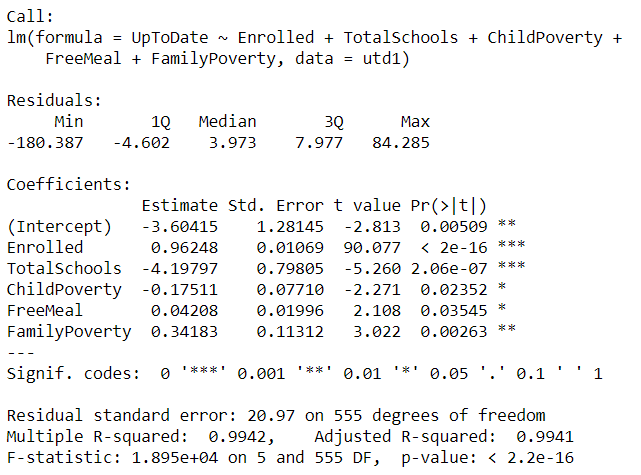
The point estimates of the Bayesian model were considerably close to the results of the frequentist approach. The Enrolled coefficient of both the Bayesian and frequentist analysis was 0.01 and the coefficient for TotalSchools was -0.98 in the Bayesian approach and -1.01 in the frequentist. Although the coefficient for Enrolled is much smaller and notably a far lesser predictor within the model, count of students enrolled and total schools showed strong support for the best predictors of district reporting completion.



*Figure 16 – Bayesian estimation of logistic regression*

### Vaccine Rate

An initial linear regression model was run using the number of students enrolled, the total schools in the district, the number of students in child poverty, the number receiving free meals, and the number in family poverty as possible predictors of the percent of students with up-to-date vaccinations. The statistically significant variables at the 0.001 level were the number of students enrolled, the total number of schools in the district. At the 0.01 level the number of students in family poverty was also a statistically significant variable in the initial model.



*Figure 17 – initial linear regression model predicting up to date students*

Removing the students receiving a free meal and those in child poverty from the model resulted in an adjusted r-square of 0.9941 with all remaining variables showing statistically significant p-values.

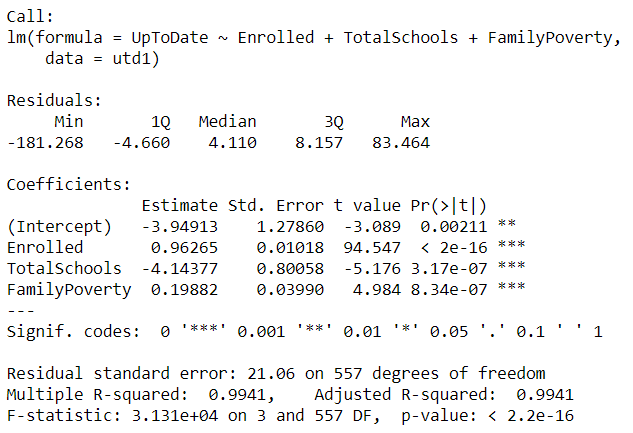
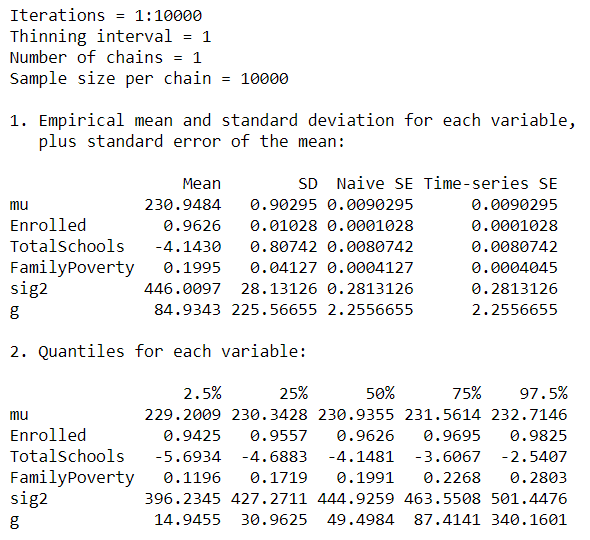


Figure 18

###### Bayesian Analysis

Additional Bayesian regression analysis was performed to further investigate the best predictors of students up to date on vaccines. The test analyzing Enrolled, TotalSchools and FamilyPoverty showed Enrolled, TotalSchools, and FamilyPoverty to have almost identical coefficients. **This indicated that the number of students enrolled, the number of schools in the district, and the number of students in family poverty were the best predictors of the percent of children with up-to-date vaccinations.**

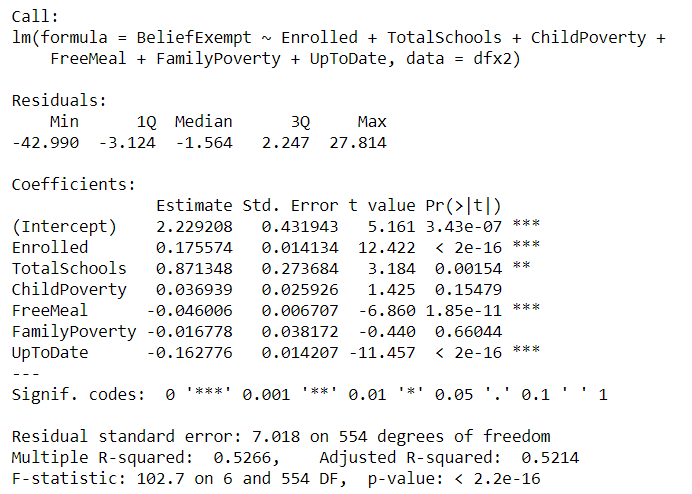
**

*Figure 19 – Bayesian regression of up-to-date student predictors*

##### Belief Exceptions

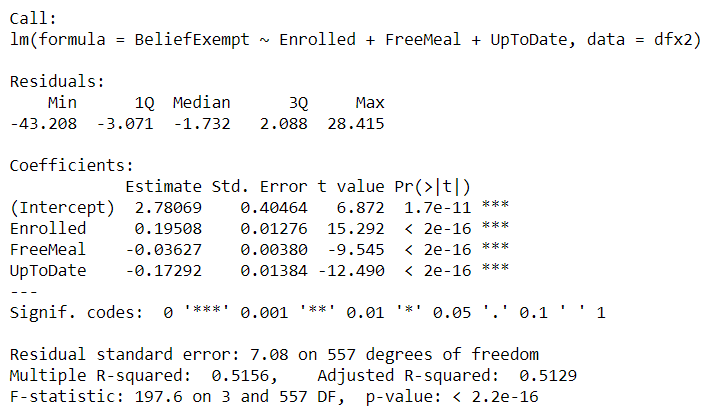
###### Frequentist Analysis

To determine the strongest predictors for the percentage of enrolled students with belief exemptions, a linear model was tested using the predictors of the number of enrolled students, the number of total schools in the district, the number of children in poverty, the number of students receiving free meals, the number of children in family poverty, and the number of students up to date on vaccinations. This initial model had an adjusted r-squared of 0.52 and showed the variables ChildPoverty, TotalSchools, and FamilyPoverty to be poor predictors in the model with p-values above the 0.001 alpha level.



*Figure 19 – initial linear regression model predicting Belief Exemptions*

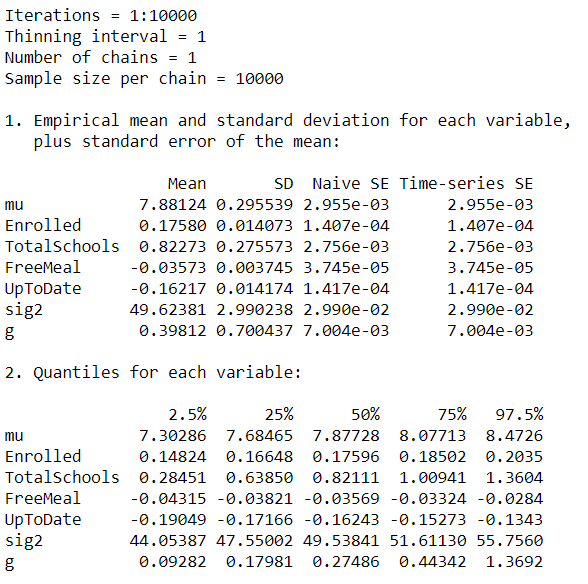
Removing the poor predictors, the regression was run again using only the number of enrolled students, the number of children receiving free meals, and the number of students with up-to-date vaccines as predictors. The updated model showed an adjusted r-squared of 0.51 and the results showed the number of enrolled students as the as the strongest predictor with a t-value of 15.29 and a p-value < 2e-16, followed by the number of students receiving a free meal as a strong negative predictor with a t-value of -9.55 and a p-value <2e-16. Finally, the number of students with up-to-date vaccinations was the third strongest predictor of students with belief exemptions with a strong negative t-value of -12.49 and a p-value <2e-16. **The number of enrolled students, the number of students receiving a free meal, and the number of students up to date on vaccinations predict the percentage of students with belief exceptions.**



*Figure 20 – updated linear regression model predicting belief exemptions*

###### Bayesian Analysis

Bayesian analysis of the same model revealed similar findings. The initial model using all possible predictors showed the 95% HDI of ChildPoverty and FamilyPoverty overlapped 0 indicating these variables were not good predictors. The revised model removing ChildPoverty and FamilyPoverty supported all remaining variables including TotalSchools. However, as the p-value of the frequentist model for TotalSchools was borderline, a Bayes Factor analysis was performed comparing the result of the model using Enrolled, FreeMeal and TotalSchools to one without TotalSchools. The model excluding TotalSchools had almost double the Bayes Factor agreeing with the frequentist approach to remove this variable from the equation. All of this strongly supports the idea that the number of students enrolled, the number of students receiving free meals and the number of students with up-to-date vaccinations are predictors for the number of students with belief exemptions.



*Figure 21 – initial Bayesian regression & HDI results*

## Conclusion

Vaccination rates increased relatively slowly from 1980 to 2017 with a notably significant drop and recovery from 1986 to 1989. Of the top 5 vaccinations in the United States during this time period the Hepatitis B vaccine had the lowest distribution rates overall but also showed the largest increase over time. The Polio vaccine experienced the greatest level of variability during the analyzed timespan with significant drops in distribution rates in the late 1980s and again in the late 2000s.

California public school reporting of vaccination data in 2013 was 11% higher than private school reporting with 97.44% and 86.36% respectively. Within the schools reporting 93% of students were vaccinated for DTP, Polio and MMR whereas 95% if students were vaccinated for HepB.

Nationally in 2013 ninety eight percent of Americans had received their first dose of the DPT vaccine. Ninety three percent had received the Pol3 & Hib3 vaccines, and 92% had received MCV1. HepB\_BD was the lowest distributed vaccine only being given to 74% of the US public. Interestingly, of the HepB\_BD vaccination data reported for public schools was at 95% which was higher than any of the other top 5 vaccinations reported by public schools in California for that year.

Analysis of students missing vaccinations demonstrated a high correlation between missing one vaccine and all others. This indicated a high probability that students who are vaccinated for one vaccine are more likely to be vaccinated for all others provided.

California district analysis revealed the best predictors of the completed prediction of a district were the number of total schools in a district followed by the number of enrolled students. The number of students enrolled, the number of schools in the district, and the number of students in family poverty were the best predictors of the percent of children with up-to-date vaccinations. The number of enrolled students, the number of students receiving a free meal, and the number of students up to date on vaccinations predict the percentage of students with belief exceptions.

## Recommendations

Recommendations to the legislature for improving district reporting compliance are to make time allowances for districts with a larger number of total schools, reduce the total number of schools for each district, or perform additional analysis on potential roadblocks in the reporting completion process for larger districts. It is also recommended to focus on incentivizing private school reporting compliance as this is another area with opportunities for improvement. To improve vaccination rates, it is recommended that financial assistance be applied most strongly toward family poverty programs, free meal programs and additional efforts to increase enrollment.

# Appendix A: Q&A

## Question 1a: How have U.S. vaccination rates varied over time?

***Average Vaccination Rates:***

The average vaccination rates for the top 5 vaccines in the United States from 1980 to 2017 is displayed in Figure 1. In 1986 the average vaccination rate was 77%. There was an observable significant drop in the vaccination rates the following year, 1987, down to 53%, the lowest recorded for the top 5 vaccines within the data. In 1988 the rate rebounded to 75.4% before declining again the following year down to 64.8% at which point the rates resumes a less variable pattern moving forward. Overall, the vaccination rates did not recover to meet or exceed the pre-1987 rate until 1996.

Observing the differences over time in Figure 2 below removes the impacts of growth to provide a more objective view of the data. From 1980 to 1986 the range of difference for the average vaccine rate was between 0 to +2.6. In 1987 the vaccination rate did indeed plumet by -24, followed by a sharp rebound in 1988 to almost +23.4 and another drop in 1989 of -11.6. In 1990 the difference over time stabilized to be within -1.2 to +4.8.

Change point analysis of the variance further supports these findings. Figure 3 below shows the change point location of variability in 1989. See appendix for data results of change point analysis.

***Top 5 Vaccination Rates:***

Breaking out the data by each of the top 5 vaccines showed additional insights into the overall trend of vaccination rates. As shown in Figure 5 below, both the Hepatitis and DTP1 vaccination rate remained stable during the 1987 vaccination rate drop while there were drops in the rates of MCV1, Hib3, and Pol3. In 1988 vaccination rates for these 3 vaccines returned to ranges close to the before the drop and then declined slightly less dramatically again in 1989. Interestingly the HepB\_BD vaccine, which had not previously been impacted by the sharp decline, suffered a decline in rates while other vaccine rates were recovering. DTP1 followed with a smaller decline in vaccination rates in 1989.

## Question 1b: Are vaccination rates increasing or decreasing?

The mean change point of the average vaccine rate was location 23 as shown in Figure 4 above. This correlates to a transition from the mean of 1980 – 2002 to a new mean from 2003 – 2017. The data suggests the jump in vaccination rates in 2003 was the start to a new average range for inoculation rates. See appendix for data results of mean change point analysis. However, referring back to Figure 3 it is important to observe that since 2012 the range of difference in vaccination rates narrowed to -1.4 to + 0.8, with 9 years having less than 1% change in vaccination rates. This indicates very small growth over time.

## Questions 1c: Which vaccination has the highest rate at the conclusion of the time series?

As can be seen [here](#Top5VaccRates), the DTP1 vaccine has the highest rate in 2017

## Question 1d: Which vaccination has the lowest rate at the conclusion of the time series?

As can be seen [here](#Top5VaccRates), the HepB\_BD has the lowest rate in 2017.

## Question 1e: Which vaccine has the greatest volatility?

Looking at the differences in each of the top 5 vaccine rates over time in Figure 6 provided insight into the ranges of change over time that impacted the average. Individual vaccine difference ranges throughout the analyzed period are reported in Table 1 below showing the overall spread of variance for distribution of each vaccine. Note that **Pol3 has the widest variability range of approximately -73 to +73 and therefore the greatest volatility.**

## Questions 2a: What proportion of public schools reported vaccination data?

Of the public schools, (0.76/0.78) 97.44% reported and (0.02/0.78) 2.56% did NOT.

## Question 2b: What proportion of private schools reported vaccination data?

Of the private schools, (0.19/0.22) 86.36% reported and (0.03/0.22) 13.64% did NOT.

## Question 2c: Was there any credible difference in overall reporting proportions between public and private schools?

Between public and private schools there was an 11% difference in reporting rates.

## 

## Question 3a: What are 2013 vaccination rates for individual vaccines (i.e., DOT, Polio, MMR, and HepB) in California public schools?

The total California public school students in 2013 was reported as 443,046. Of these, 93% had their DTP, Polio & MMR vaccines and 7% did not. 95% of children enrolled in public school in California in 2013 had their HepB vaccine and 5% did not.

## Question 3b: 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)?

Nationally in 2013 ninety eight percent of Americans had received their first dose of the DPT vaccine. Ninety three percent had received the Pol3 & Hib3 vaccines, and 92% had received MCV1. HepB\_BD was the lowest distributed vaccine only being given to 74% of the US public. Interestingly, of the HepB\_BD vaccination data reported for public schools was at 95% which was higher than any of the other top 5 vaccinations reported by public schools in California for that year.

## Question 4: 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?

Initial correlation analysis (above) shows high positive correlation between all 4 of the variables WithoutDTP, WithoutPolio, WithoutMMR, and WithoutHepB. To further illustrate these correlations, additional Frequentist PPMC and Bayesian Correlation tests were executed between WithoutDTP, WithoutPolio, WithoutMMR & WithoutHepB. **The strength of these correlations indicates there is a high correlation between students who are missing one vaccine and missing other vaccines.** The frequentist and Bayesian analysis both support the idea that students missing one vaccine in the data are far more likely to be missing another vaccine.

## Question 5: What variables predict whether or not a district’s reporting was complete?

California district analysis revealed the best predictors of the completed prediction of a district were the number of total schools in a district followed by the number of enrolled students.

## Question 6: What variables predict the percentage of all enrolled students with completely up-to-date vaccines?

The number of students enrolled, the number of schools in the district, and the number of students in family poverty were the best predictors of the percent of children with up-to-date vaccinations.

## Question 7: What variables predict the percentage of all enrolled students with belief exceptions?

The number of enrolled students, the number of students receiving a free meal, and the number of students up to date on vaccinations predict the percentage of students with belief exceptions.

## Question 8: What’s the big picture, based on all of the foregoing analyses? The staff member in the state legislator’s 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?

## Conclusion

Vaccination rates increased relatively slowly from 1980 to 2017 with a notably significant drop and recovery from 1986 to 1989. Of the top 5 vaccinations in the United States during this time period the Hepatitis B vaccine had the lowest distribution rates overall but also showed the largest increase over time. The Polio vaccine experienced the greatest level of variability during the analyzed timespan with significant drops in distribution rates in the late 1980s and again in the late 2000s.

California public school reporting of vaccination data in 2013 was 11% higher than private school reporting with 97.44% and 86.36% respectively. Within the schools reporting 93% of students were vaccinated for DTP, Polio and MMR whereas 95% if students were vaccinated for HepB.

Nationally in 2013 ninety eight percent of Americans had received their first dose of the DPT vaccine. Ninety three percent had received the Pol3 & Hib3 vaccines, and 92% had received MCV1. HepB\_BD was the lowest distributed vaccine only being given to 74% of the US public. Interestingly, of the HepB\_BD vaccination data reported for public schools was at 95% which was higher than any of the other top 5 vaccinations reported by public schools in California for that year.

Analysis of students missing vaccinations demonstrated a high correlation between missing one vaccine and all others. This indicated a high probability that students who are vaccinated for one vaccine are more likely to be vaccinated for all others provided.

California district analysis revealed the best predictors of the completed prediction of a district were the number of total schools in a district followed by the number of enrolled students. The number of students enrolled, the number of schools in the district, and the number of students in family poverty were the best predictors of the percent of children with up-to-date vaccinations. The number of enrolled students, the number of students receiving a free meal, and the number of students up to date on vaccinations predict the percentage of students with belief exceptions.

## Recommendations

Recommendations to the legislature for improving district reporting compliance are to make time allowances for districts with a larger number of total schools, reduce the total number of schools for each district, or perform additional analysis on potential roadblocks in the reporting completion process for larger districts. It is also recommended to focus on incentivizing private school reporting compliance as this is another area with opportunities for improvement. To improve vaccination rates, it is recommended that financial assistance be applied most strongly toward family poverty programs, free meal programs and additional efforts to increase enrollment.