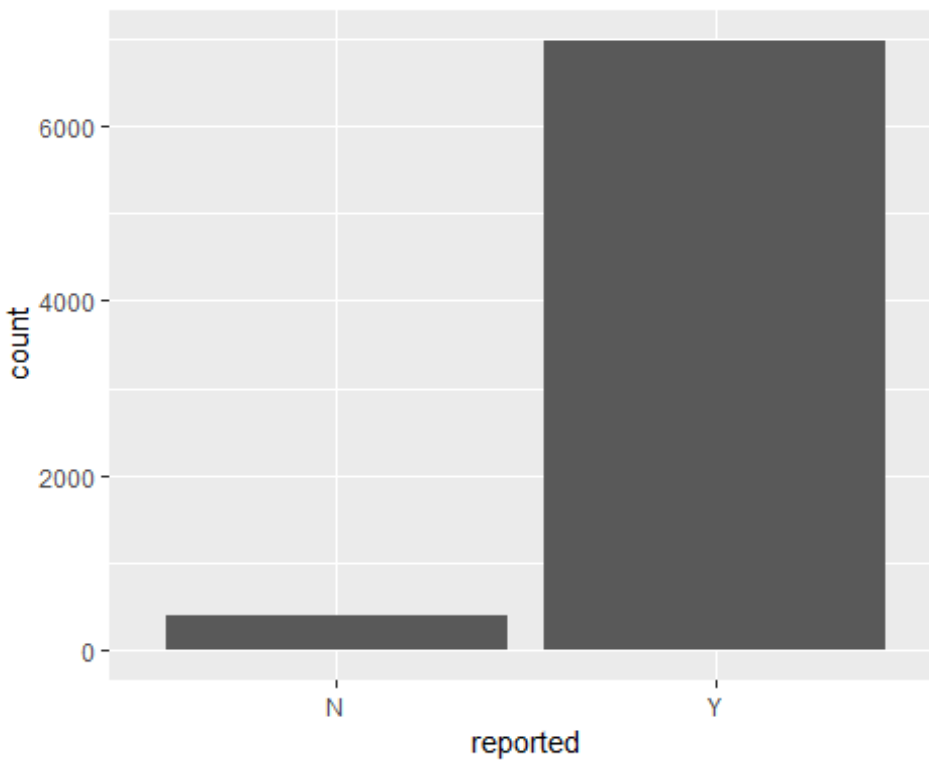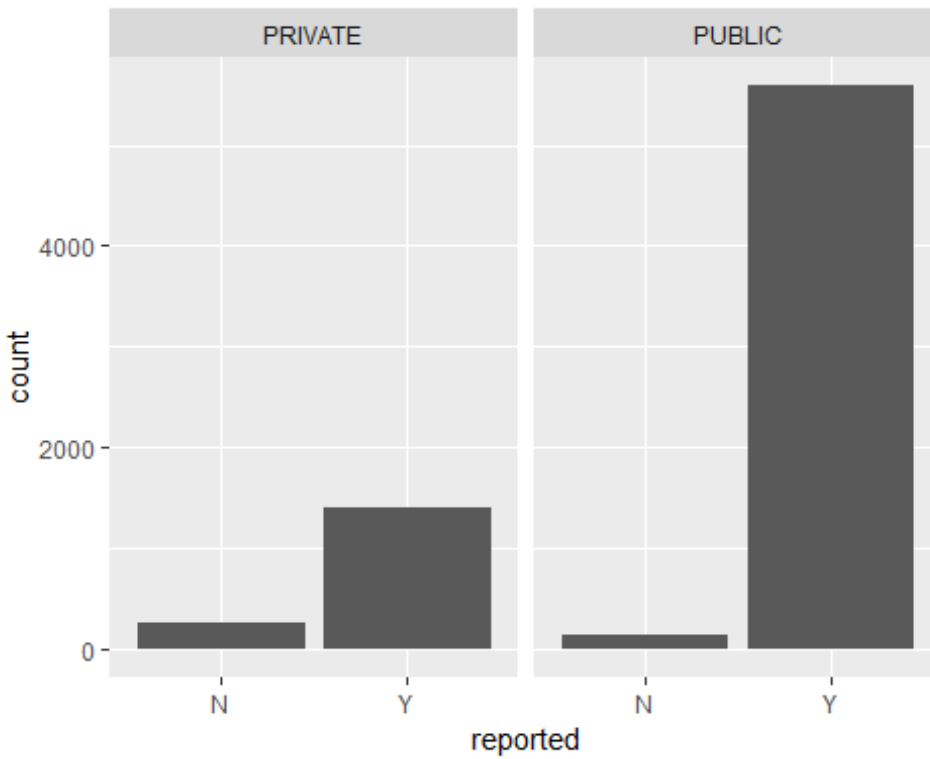# IST 777: final project

Jeff Levesque

Introductory / Descriptive Reports

Proportion of overall schools reported vaccination data

Figure 1: histogram signifying overall proportion reporting vaccinations. See Appendix A for the associated code.

Figure 2: histogram signifying proportion of public and private schools reporting vaccinations. See Appendix B for the associated code.

	Private	Public
N	252	148
Y	1397	5584

Figure 3: tabulated counts of schools reporting vaccinations. See Appendix C for the associated code.

|	Private	Public
N	0.03414172	0.02005148
Y	0.18926975	0.75653705

Figure 4: normalized counts of schools reporting vaccinations. See Appendix D for the associated code.

A significant proportion of public schools have reported more vaccinations than private school. Moreover, both schools who have not reported vaccinations, represent a very small marginal total.

U.S. vaccinations rates over time

Most vaccinations have stabilized. However, it appears that `HepB_BD` vacinations started much recently, and may require additional time to stabilize:
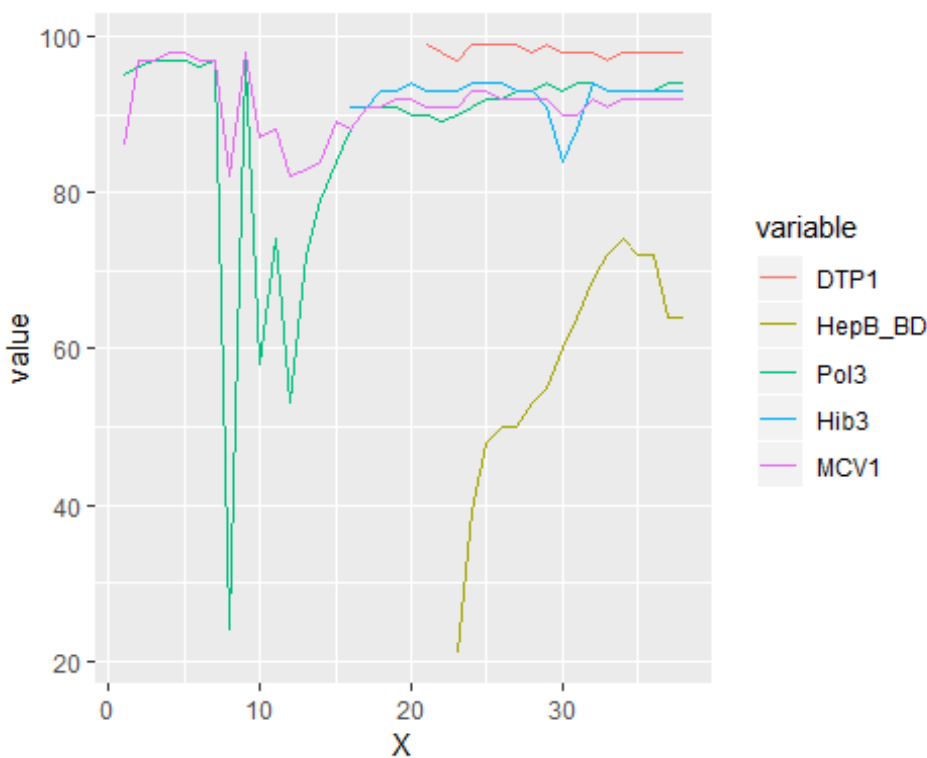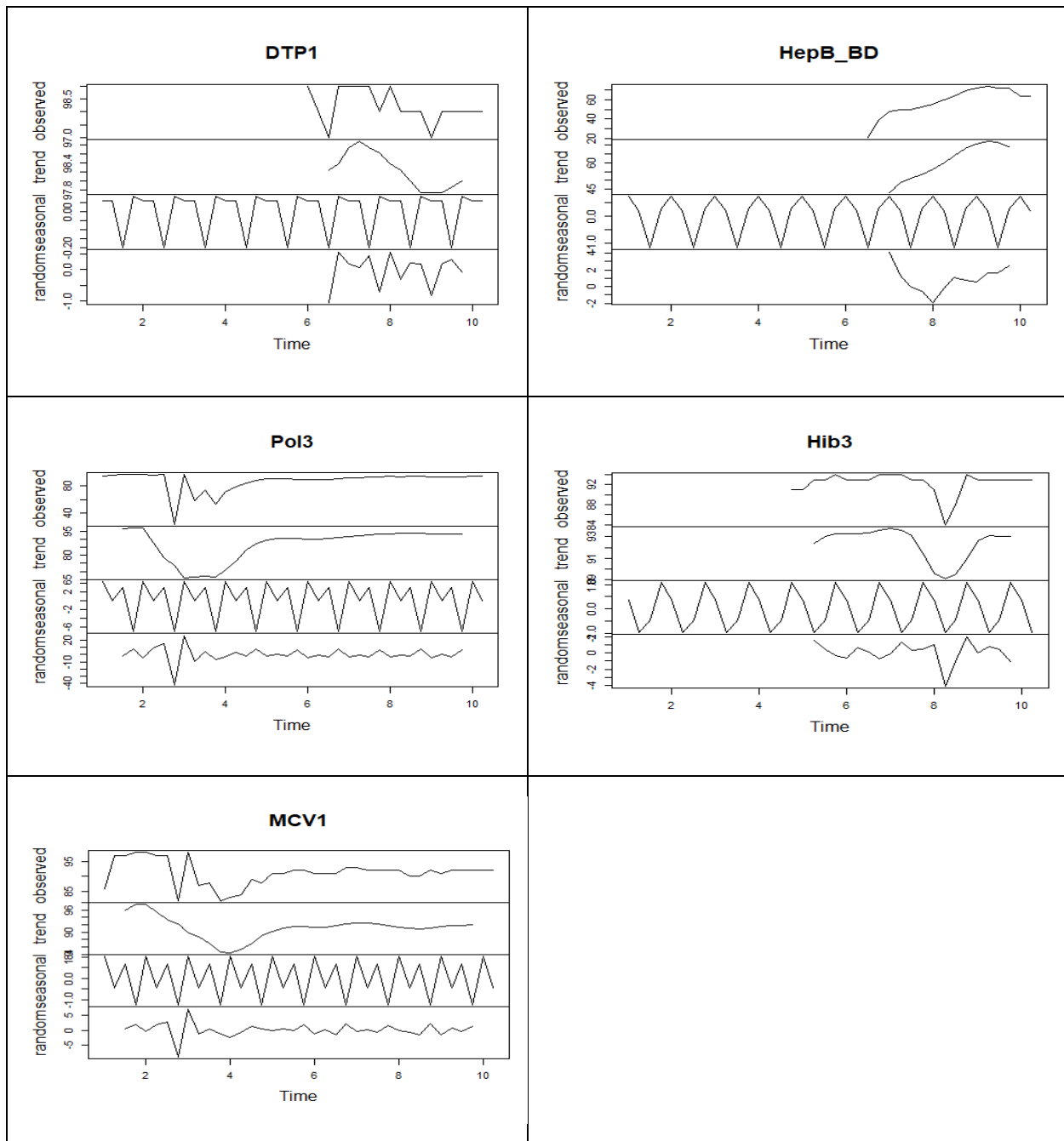

Figure 5: time series plots for all vaccination types. See Appendix E for associated code.

In general the vacination rates reported by the World Health Organization for the United States, have a tendency to fluctuate initially. After some time, the rates seem to stabilize between 85 and 100. Furthermore, the `HepB_BD` rates does not show similar values, nor stability. Rather, the vacination endured a large growth period, then a decline. It is unknown whether the rates have stabilized, or future volatility could occur.

Notable patterns in U.S. vaccinations rates over time

When using time series decomposition, with a `frequency=4`, there appears to be seasonal patterns for all vacination types. However, the trends are not consistent nor similar between the different vacination types.

Figure 6: decomposed time series plots for all vaccination types. See Appendix F for associated code.

Public vs. Private School Comparisons

Difference in overall proportions between public and private schools
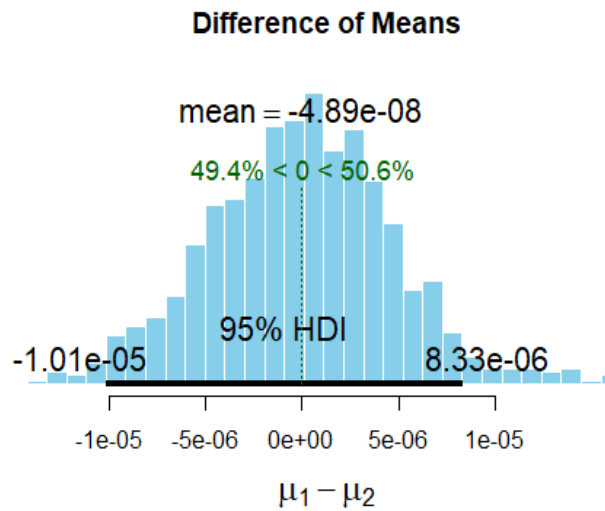
Frequentist t-test

The frequentist null hypothesis states there are no difference in proportions between the public and private schools. The associated alternative hypothesis, suggests there exists a significant difference in proportions. Using the Welch two sample t-test, the corresponding critical value 2.2e-16 < 0.05 = alpha value. This signifies a significant difference in proportions between public and private school.

```
t.test(
  x=allSchoolsReportStatus[allSchoolsReportStatus$pubpriv=='PUBLIC', 1],
  y=allSchoolsReportStatus[allSchoolsReportStatus$pubpriv=='PRIVATE', 1]
)

##
##  Welch Two Sample t-test
##
## data:  allSchoolsReportStatus[allSchoolsReportStatus$pubpriv == "PUBLIC",
and allSchoolsReportStatus[allSchoolsReportStatus$pubpriv == "PRIVATE",     1
] and      1]
## t = -148.78, df = 7379, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3739.090 -3641.841
## sample estimates:
## mean of x mean of y
##  2866.508  6556.973
```

Bayes Factor (with HDI)

The Bayesian HDI is a 95% density region, where points within this region are more credible than the remaining 5% distributed among the tails. Specifically, the interval spans (-8.99e-06, 9.18e-06). Since this interval overlaps zero, there no credible evidence suggesting a mean difference between the report status of public and private school.

Figure 7: high density interval for overall proportions. See Appendix G for associated code.

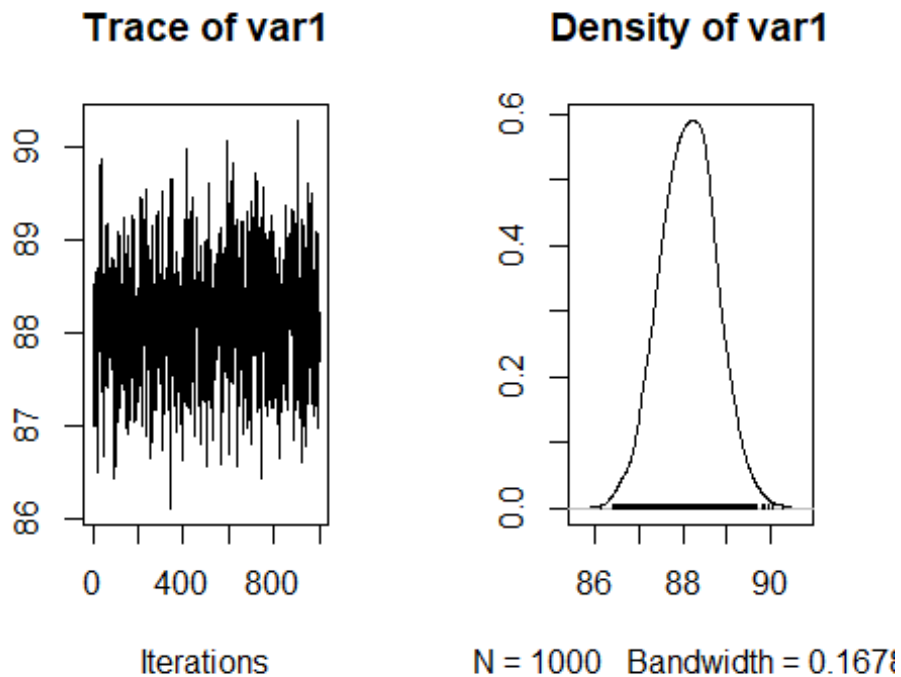## Overall vaccination rates between public and private schools

Frequentist: t-test

The frequentist null hypothesis states there are no difference in proportions between the overall vaccination rates between public and private schools. The associated alternative hypothesis, suggests there exists a significant difference in proportions. Using the Welch two sample t-test, a critical value $0.085 > 0.05$ = alpha value (see Appendix H for associated code). Therefore, there is insufficient evidence to reject the null hypothesis.

Bayes Factor

The bayes factor provides an odds ratio in favor of the alternative hypothesis, namely there is a difference in proportions between the overall vaccination rates. In the below case, the odds ratio is 0.813:1 in favor of the alternative hypothesis. This suggests an inverted 1/0.813 => 1.23:1 ratio in favor of the null hypothesis (see appendix I for associated code). Though this ratio is greater than 1, it is significantly less than 3, and not worth mentioning.

Furthermore, the corresponding traceplot visually indicates some spikes at roughly 375, and almost 900. These outliers suggest that the markov chain monte carlo (MCMC), may not have converged. Thus using higher `iterations` could produce better results. However, the associated density curve, indicates a roughly normal distribution with no outliers.

Figure 8: trace and density plot for overall vaccination rates. See Appendix I for associated code.

HDI

The Bayesian HDI is a 95% density region, where points within this region are more credible than the remaining 5% distributed among the tails. Specifically, the interval spans (0, 200). Since this interval includes zero, there no credible evidence suggesting a mean difference between the report status of public and private school.
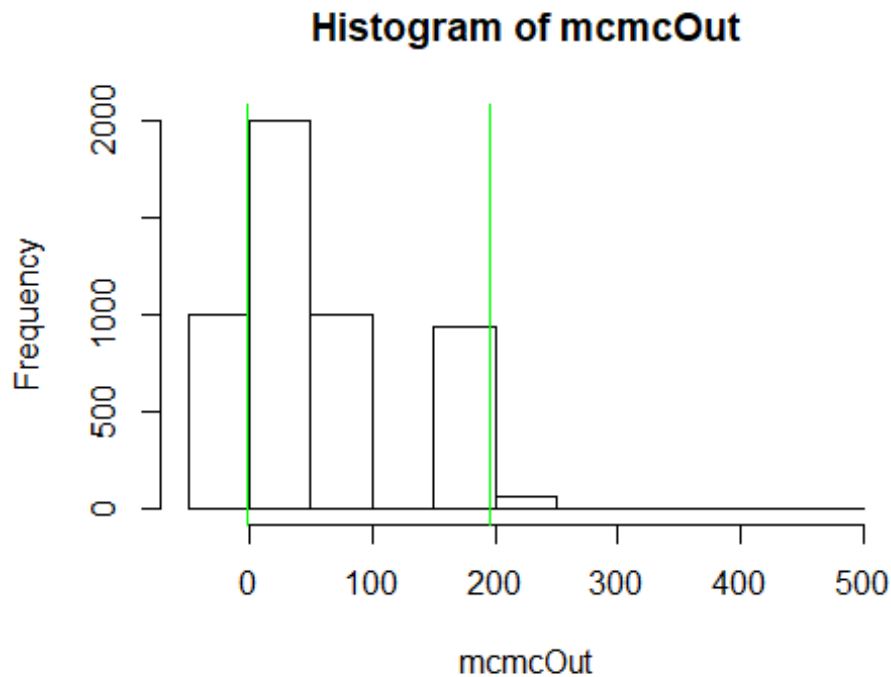
Figure 9: HDI for overall vaccination rates. See Appendix J for associated code.

Medical exemptions between public and private schools

Frequentist: t-test

The frequentist null hypothesis states there are no difference in proportions of medical exemptions between public and private schools. The associated alternative hypothesis, suggests there exists a significant difference in proportions. Using the Welch two sample t-test, a critical value 0.2774 > 0.05 = alpha value (see Appendix K for associated code). Therefore, there is insufficient evidence to reject the null hypothesis.

Bayes Factor

The bayes factor provides an odds ratio in favor of the alternative hypothesis, namely there is a difference in proportions of medical exemptions between public and private schools. In the below case, the odds ratio is 0.179:1 in favor of the alternative hypothesis. This suggests an inverted 1/0.179 => 5.5:1 ratio in favor of the null hypothesis (see Appendix L for associated code).

Furthermore, the corresponding traceplot visually indicates some spikes at roughly 400, 700, and almost 900. These outliers suggest that the markov chain monte carlo (MCMC), may not have converged. Thus using higher `iterations` could produce better results.

However, the associated density curve indicates a mostly normal distribution, while the central mode area curving slightly.

Trace of var1   **Density of var1**

N = 1000 Bandwidth = 0.0245
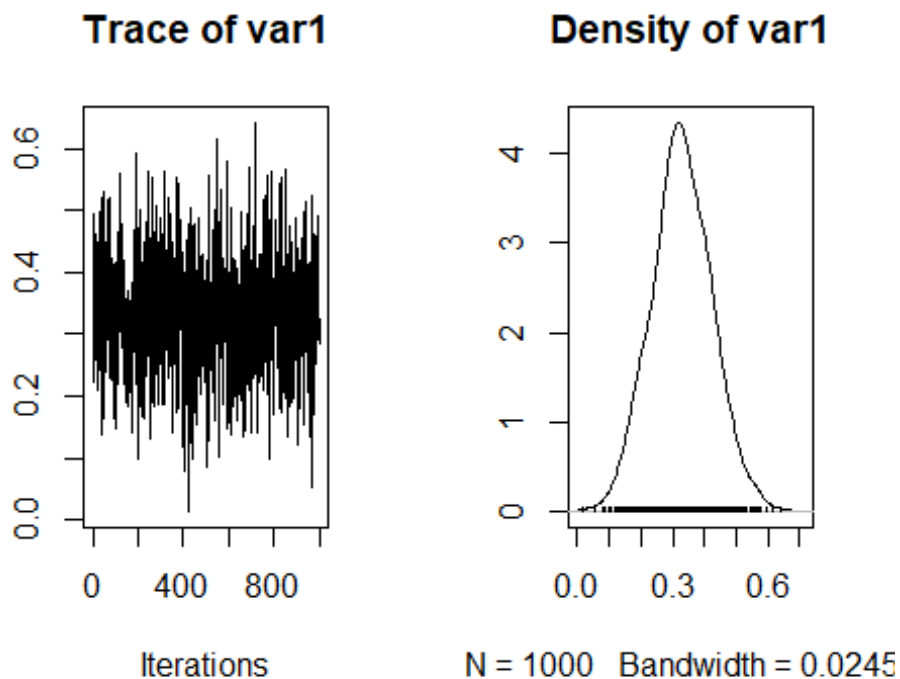
Figure 10: trace and density plot for medical exemptions between public and private schools. See Appendix L for associated code.

HDI

The Bayesian HDI is a 95% density region, where points within this region are more credible than the remaining 5% distributed among the tails. In the below diagram, the interval includes zero, suggesting no credible evidence for a mean difference of medical exemptions between public and private schools.

Histogram of mcmcOut
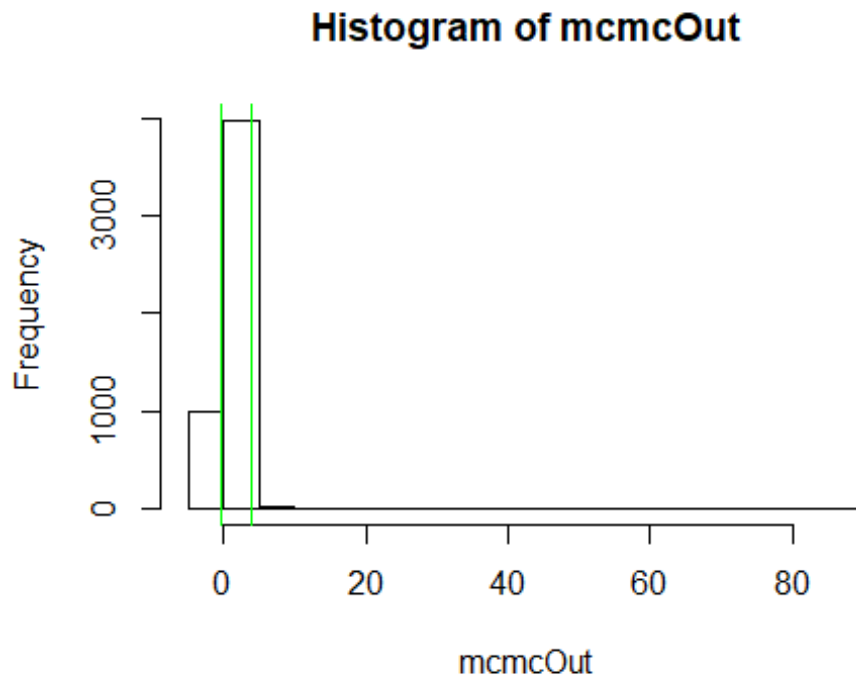
Figure 10: HDI for medical exemptions between public and private schools. See Appendix M for associated code.

Religious/belief exemptions between public and private schools

Frequentist: t-test

The frequentist null hypothesis states there are no difference in proportions of religious/belief exemptions between public and private schools. The associated alternative hypothesis, suggests there exists a significant difference in proportions. Using the Welch two sample t-test, a critical value $0.125 > 0.05$ = alpha value (see Appendix N for associated code). Therefore, there is insufficient evidence to reject the null hypothesis.

Bayes Factor

The bayes factor provides an odds ratio in favor of the alternative hypothesis, namely there is a difference in proportions of religious/belief exemptions between public and private schools. In the below case, the odds ratio is 0.442:1 in favor of the alternative hypothesis. This suggests an inverted 1/0.442 => 2.262:1 ratio in favor of the null hypothesis. Though this ratio is greater than 1, it is less than 3, and not worth mentioning (see Appendix O for associated code).

Furthermore, the corresponding traceplot visually indicates some spikes at roughly 200, 250, 600, 900. These outliers suggest that the markov chain monte carlo (MCMC), may not

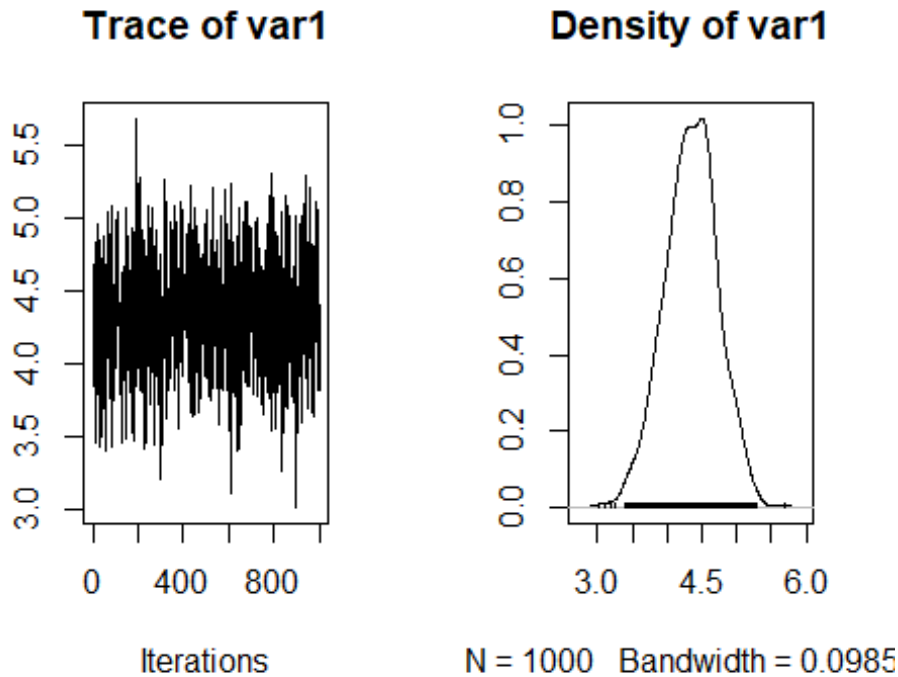have converged. Thus using higher `iterations` could produce better results. Furthermore, the associated density curve indicates a mostly normal distribution with a slighlty distorted peak.

Figure 11: trace and density plot for religious/belief exemptions between public and private schools. See Appendix O for associated code.

HDI

The Bayesian HDI is a 95% density region, where points within this region are more credible than the remaining 5% distributed among the tails. In the below diagram, the interval includes zero, suggesting no credible evidence for a mean difference of religious/belief exemptions between public and private schools.

Histogram of mcmcOut
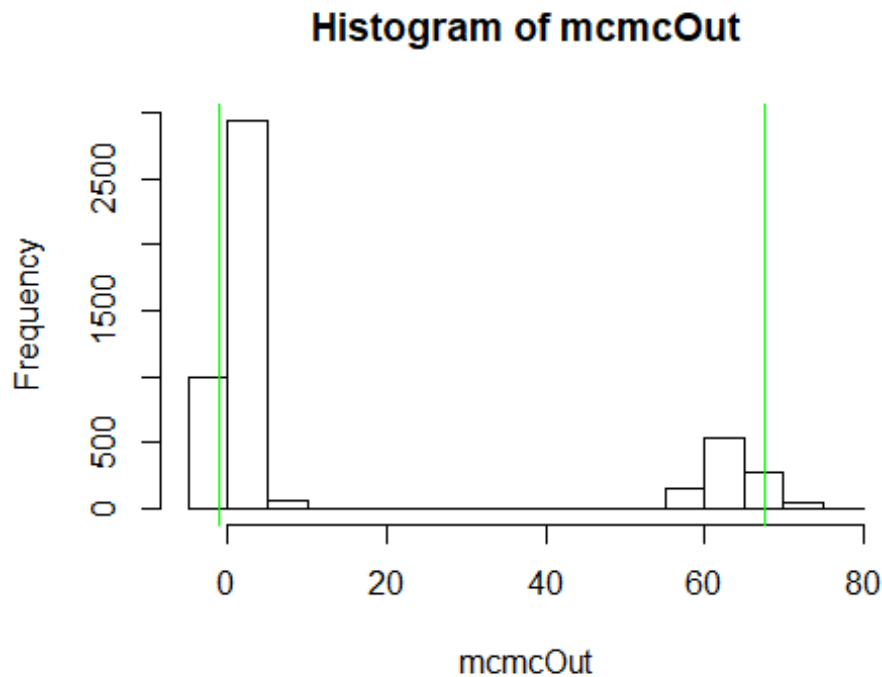
Figure 12: HDI for religious/belief exemptions between public and private schools. See Appendix P for associated code.

Predictive Analyses

Predict public or private

Predicting a bivariate condition is possible using the logistic regression. By ensuring the dependent variable is a `factor`, and the `glm` function can be implemented (see Appendix Q for associated code). The resulting logistic model has the following log odds results:

- for every unit change in conditional, the log odds of predicting a public school reporting vaccination changes by `-0.009361`.
- for every unit change in medical, the log odds of predicting a public school reporting vaccination changes by `-0.035439`.
- for every unit change in religious, the log odds of predicting a public school reporting vaccination changes by `0.0993`.

Moreover, each independent variable is found not significant. The religious independent variable had the lowest with `0.0993 > 0.05` = alpha level.

Additionally, the residuals vs. fitted plot indicate that public school are being overpredicted. This signifies highly positive residuals (as indicated above `666.6`), and

generally a poor logistic regression. The corresponding Normal Q-Q plot does not provide much information other than ressembling the logistic pattern.
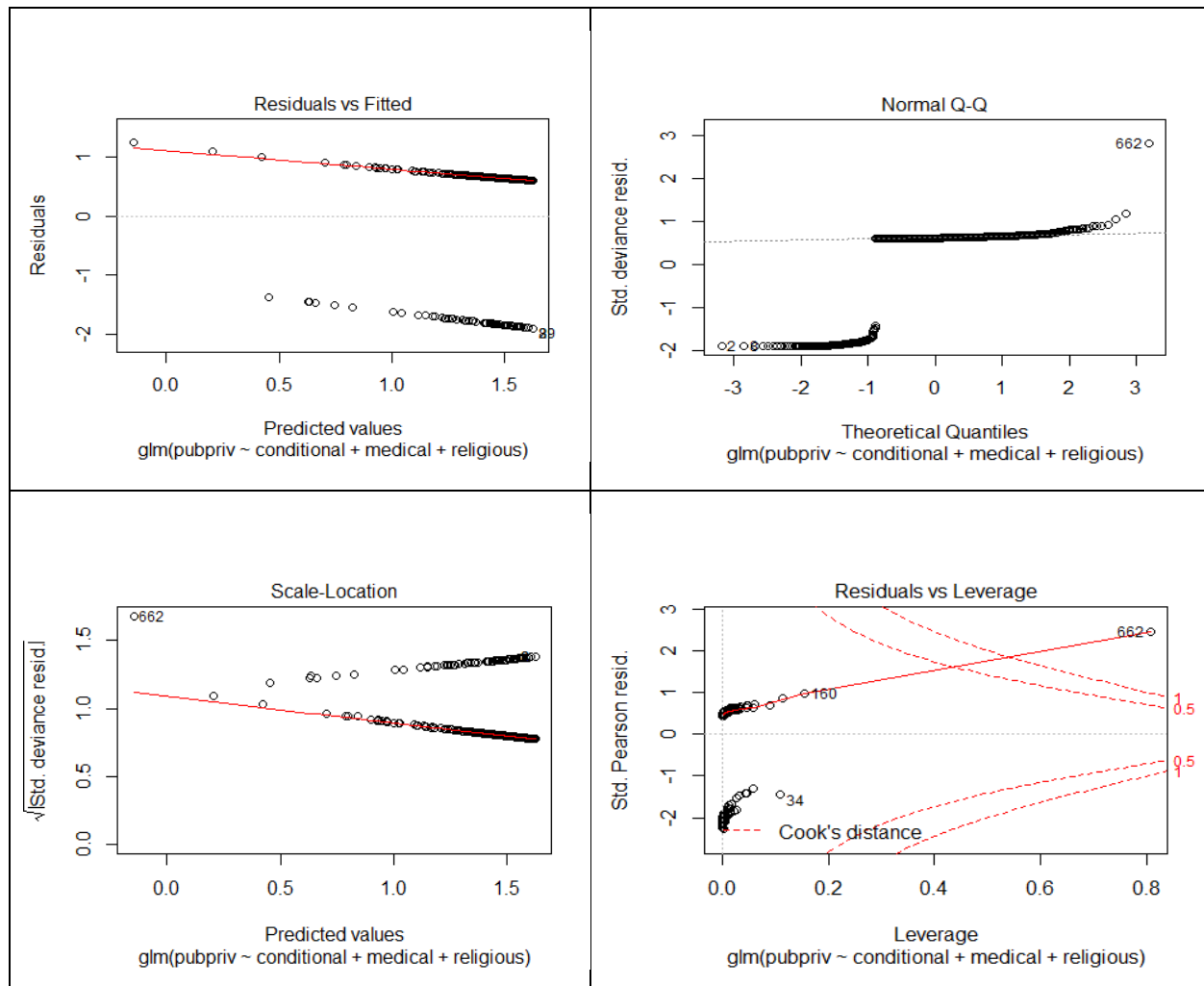
Figure 13: Logistic regression diagnostic. See Appendix Q for associated code.

Predict conditional percentage

The following linear regression indicates that each independent variable, except `hepMiss` is statistically significant at an alpha = 0.05. Moreover, the general p-value = `2.2e-16` suggests that the overall model is statistically significant:

```
conditional = lm(
  conditional ~ dptMiss + polMiss + mmrMiss + hepMiss + varMiss,
  data = reportSample
```

```
)
summary(conditional)

##
## Call:
## lm(formula = conditional ~ dptMiss + polMiss + mmrMiss + hepMiss +
##     varMiss, data = reportSample)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -22.391  -1.923  -0.880   0.724  35.453
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.88022    0.27239   3.231 0.001290 **
## dptMiss      0.68447    0.10038   6.819 2.00e-11 ***
## polMiss      0.42397    0.10893   3.892 0.000109 ***
## mmrMiss      0.34747    0.07188   4.834 1.65e-06 ***
## hepMiss     -0.20081    0.10502  -1.912 0.056262 .
## varMiss     -0.99670    0.09830 -10.140  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.678 on 692 degrees of freedom
## Multiple R-squared:  0.7227, Adjusted R-squared:  0.7207
## F-statistic: 360.6 on 5 and 692 DF,  p-value: < 2.2e-16
```

When running the equivalent bayes factor, the posterior HDI indicates only hepMiss (-0.4065, 0.00463) overlaps zero. This indicates that the B-weights credibly differ from zero:

```
bf.conditional = lmBF(
  data = reportSample,
  conditional ~ dptMiss + polMiss + mmrMiss + hepMiss + varMiss,
  posterior = TRUE,
  iterations = 1000
)
summary(bf.conditional)

##
## Iterations = 1:1000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 1000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##             Mean      SD Naive SE Time-series SE
## mu       6.8731 0.21489 0.006796       0.006513
## dptMiss  0.6797 0.09953 0.003147       0.003147
## polMiss  0.4262 0.10563 0.003340       0.003157
```

```
## mmrMiss   0.3465 0.07029 0.002223        0.002223
## hepMiss  -0.1978 0.10451 0.003305        0.003305
## varMiss  -0.9968 0.09899 0.003130        0.002824
## sig2      32.4438 1.70930 0.054053        0.054053
## g          0.6458 0.55392 0.017516        0.017516
##
## 2. Quantiles for each variable:
##
##               2.5%      25%     50%      75%      97.5%
## mu          6.4336   6.7275   6.8723   7.0259   7.270097
## dptMiss     0.4861   0.6121   0.6802   0.7479   0.864083
## polMiss     0.2242   0.3577   0.4259   0.4962   0.624085
## mmrMiss     0.2097   0.2974   0.3458   0.3939   0.484008
## hepMiss    -0.4074  -0.2661  -0.1963  -0.1286   0.000703
## varMiss    -1.1955  -1.0615  -0.9979  -0.9314  -0.808865
## sig2       29.2177  31.3098  32.3332  33.5197  35.904171
## g           0.1933   0.3419   0.4959   0.7654   1.998237
```

Furthemore, the bayes r-squared 0.718931 is less than 0.7227 from the conventional analysis:

```
rsquared = 1 - (bf.conditional[, 'sig2'] / var(reportSample$conditional))
mean(rsquared)
```

```
## [1] 0.7188977
```

```
quantile(rsquared, c(0.025))
```

```
##      2.5%
## 0.6889158
```

```
quantile(rsquared, c(0.975))
```

```
##     97.5%
## 0.746849
```

Both approaches indicate a fairly strong relationship between the dependent variable (conditional), with the associated independent variables.

In an attempt to investigate the above coefficients, a chi-squared test was implemented as an omnibus test. The corresponding chi-squared test shows that polMiss, and mmrMiss are not statistically significant:

```
anova(conditional, test='Chisq')
```

```
## Analysis of Variance Table
##
## Response: conditional
##             Df Sum Sq Mean Sq    F value Pr(>F)
## dptMiss      1  35839   35839 1111.5795 <2e-16 ***
## polMiss      1     90      90    2.8003 0.0947 .
## mmrMiss      1    107     107    3.3051 0.0695 .
```

```
## hepMiss      1  18784   18784  582.5946 <2e-16 ***
## varMiss      1   3315    3315  102.8102 <2e-16 ***
## Residuals 692  22311      32
## ---
## Signif. codes:   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Predict medical percentage

The following linear regression indicates that each independent variable, except `varMiss` is not statistically significant at an alpha = 0.05. Moreover, the general p-value = `1.995e-10` suggests that the overall model is statistically significant:

```
conditional = lm(
  medical ~ dptMiss + polMiss + mmrMiss + hepMiss + varMiss,
  data = reportSample
)
summary(conditional)

##
## Call:
## lm(formula = medical ~ dptMiss + polMiss + mmrMiss + hepMiss +
##      varMiss, data = reportSample)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.050 -0.275 -0.060  0.011 41.685
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.01134    0.08991  -0.126   0.8997
## dptMiss      0.01090    0.03313   0.329   0.7422
## polMiss     -0.04336    0.03596  -1.206   0.2283
## mmrMiss      0.02103    0.02373   0.886   0.3757
## hepMiss     -0.01663    0.03466  -0.480   0.6315
## varMiss      0.08350    0.03245   2.573   0.0103 *
## ---
## Signif. codes:   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.874 on 692 degrees of freedom
## Multiple R-squared:  0.07505,    Adjusted R-squared:  0.06837
## F-statistic: 11.23 on 5 and 692 DF,  p-value: 1.995e-10
```

When running the equivalent bayes factor, the posterior HDI indicates only `hepMiss` (0.01741, 0.14185) does not overlap zero. This indicates a majority of variables have B-weights that do no credibly differ from zero:

```
bf.conditional = lmBF(
  data = reportSample,
  medical ~ dptMiss + polMiss + mmrMiss + hepMiss + varMiss,
  posterior = TRUE,
```

```
   iterations = 1000
)
summary(bf.conditional)

##
## Iterations = 1:1000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 1000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##              Mean      SD  Naive SE Time-series SE
## mu        0.27047 0.06836 0.0021617      0.0021617
## dptMiss   0.01066 0.03267 0.0010332      0.0010332
## polMiss  -0.04214 0.03440 0.0010877      0.0011979
## mmrMiss   0.02113 0.02333 0.0007376      0.0007376
## hepMiss  -0.01582 0.03426 0.0010834      0.0010834
## varMiss   0.07948 0.03190 0.0010087      0.0010087
## sig2      3.49891 0.18209 0.0057582      0.0057582
## g         0.04942 0.04541 0.0014359      0.0014359
##
## 2. Quantiles for each variable:
##
##               2.5%       25%       50%       75%   97.5%
## mu         0.13830  0.225432  0.27146  0.316447 0.40708
## dptMiss   -0.05040 -0.011099  0.01065  0.032471 0.07354
## polMiss   -0.10892 -0.065049 -0.04215 -0.018907 0.02518
## mmrMiss   -0.02186  0.005897  0.02119  0.037234 0.06759
## hepMiss   -0.08591 -0.038388 -0.01623  0.007188 0.05102
## varMiss    0.01579  0.058993  0.07867  0.100635 0.14202
## sig2       3.16368  3.371473  3.49708  3.616435 3.87362
## g          0.01452  0.025309  0.03686  0.056080 0.17870
```

Furthemore, the bayes r-squared 0.07022636 is smaller than 0.07505 from the conventional analysis:

```
rsquared = 1 - (bf.conditional[, 'sig2'] / var(reportSample$medical))
mean(rsquared)

## [1] 0.07206285

quantile(rsquared, c(0.025))

##        2.5%
## -0.02731174

quantile(rsquared, c(0.975))
```

```
##      97.5%
## 0.1609694
```

However, both approaches indicate that the model does not predict a relationship between the dependent variable (medical), with the associated independent variables.

In an attempt to investigate the above coefficients, a chi-squared test was implemented as an omnibus test. The corresponding chi-squared test shows that `polMiss`, and `mmrMiss` are not statistically significant:

```
anova(conditional, test='Chisq')

## Analysis of Variance Table
##
## Response: medical
##            Df  Sum Sq Mean Sq F value     Pr(>F)
## dptMiss     1  109.84 109.843 31.2690 3.235e-08 ***
## polMiss     1    0.58   0.578  0.1646   0.68508
## mmrMiss     1    7.40   7.405  2.1079   0.14700
## hepMiss     1   56.15  56.155 15.9856 7.069e-05 ***
## varMiss     1   23.26  23.264  6.6224   0.01028 *
## Residuals 692 2430.89   3.513
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Predict religious percentage

The following linear regression indicates `polMiss`, `mmrMiss`, and the corresponding intercept is not statistically significant at an alpha = 0.05. Moreover, the general p-value = `2.2e-16` suggests that the overall model is statistically significant:

```
conditional = lm(
  religious ~ dptMiss + polMiss + mmrMiss + hepMiss + varMiss,
  data = reportSample
)
summary(conditional)

##
## Call:
## lm(formula = religious ~ dptMiss + polMiss + mmrMiss + hepMiss +
##     varMiss, data = reportSample)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -38.879  -0.418   0.019   0.979  19.228
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.25928    0.18838   1.376 0.169153
## dptMiss     -0.24552    0.06942  -3.537 0.000432 ***
## polMiss      0.11264    0.07533   1.495 0.135313
```

```
## mmrMiss      -0.02108      0.04971   -0.424 0.671685
## hepMiss       0.27997      0.07263    3.855 0.000127 ***
## varMiss       0.60099      0.06798    8.841  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.927 on 692 degrees of freedom
## Multiple R-squared:  0.7599, Adjusted R-squared:  0.7582
## F-statistic: 438.1 on 5 and 692 DF,  p-value: < 2.2e-16
```

When running the equivalent bayes factor, the posterior HDI indicates only polMiss (-0.03557, 0.2612), mmrMiss (-0.11431, 0.0825) does overlap zero. This indicates that the B-weights credibly differ from zero:

```
bf.conditional = lmBF(
  data = reportSample,
  religious ~ dptMiss + polMiss + mmrMiss + hepMiss + varMiss,
  posterior = TRUE,
  iterations = 1000
)
summary(bf.conditional)

##
## Iterations = 1:1000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 1000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##               Mean      SD Naive SE Time-series SE
## mu       3.93755 0.14217 0.004496       0.004496
## dptMiss -0.24075 0.06937 0.002194       0.002194
## polMiss  0.11145 0.07561 0.002391       0.002391
## mmrMiss -0.02357 0.05199 0.001644       0.001712
## hepMiss  0.27824 0.07265 0.002297       0.002297
## varMiss  0.60034 0.06900 0.002182       0.002182
## sig2    15.47687 0.81877 0.025892       0.025892
## g        0.81405 0.68700 0.021725       0.022771
##
## 2. Quantiles for each variable:
##
##              2.5%     25%     50%     75%    97.5%
## mu        3.64530  3.8469  3.9382  4.0338  4.21854
## dptMiss -0.37637 -0.2872 -0.2390 -0.1962 -0.10623
## polMiss -0.02847  0.0588  0.1086  0.1600  0.26534
## mmrMiss -0.12301 -0.0602 -0.0242  0.0118  0.07948
## hepMiss  0.14438  0.2269  0.2752  0.3268  0.41779
## varMiss  0.45774  0.5555  0.6024  0.6470  0.72923
```

```
## sig2     13.95931 14.9073 15.4716 15.9863 17.27164
## g          0.22095  0.4246  0.6282  0.9733  2.50443
```

Furthemore, the bayes r-squared `0.7571835` is smaller than `0.7599` from the conventional analysis:

```
rsquared = 1 - (bf.conditional[, 'sig2'] / var(reportSample$religious))
mean(rsquared)
```

```
## [1] 0.7572937
```

```
quantile(rsquared, c(0.025))
```

```
##      2.5%
## 0.7291484
```

```
quantile(rsquared, c(0.975))
```

```
##     97.5%
## 0.7810919
```

Both approaches indicate a fairly strong relationship between the dependent variable (religious), with the associated independent variables.

In an attempt to investigate the above coefficients, a chi-squared test was implemented as an omnibus test. The corresponding chi-squared test shows that all variables were statistically significant:

```
anova(conditional, test='Chisq')
```

```
## Analysis of Variance Table
##
## Response: religious
##             Df  Sum Sq Mean Sq    F value    Pr(>F)
## dptMiss      1 20504.3 20504.3 1329.6994 < 2.2e-16 ***
## polMiss      1  1633.8  1633.8  105.9541 < 2.2e-16 ***
## mmrMiss      1   113.9   113.9    7.3873  0.006733 **
## hepMiss      1 10318.2 10318.2  669.1364 < 2.2e-16 ***
## varMiss      1  1205.2  1205.2   78.1565 < 2.2e-16 ***
## Residuals 692 10670.8    15.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Big picture

Based on the earlier analysis, the logistic regression for predicting the `pubpriv` case on condition, was not a good model since none of the independent variables were significant. Furthemore, the corresponding diagnostic tools, such as the Residuals vs. Predicted indicated that the logistic model overpredicted.

Moreover, the percentage of students requiring a follow-up on vaccination records (i.e. conditional), were statistically significant with all factors, except `hepMiss`. Hepatitis

vaccination is generally recommended by the CDC to infants between 6-23 months of age. Once a child has been provided treatment, this vaccine generally lasts a minimum of 10 years (hepA), and 20 years (hepB). This suggests if an infant receives proper vaccination shortly after birth, hepatitis should not be a problem through elementary school years.

Students with a religious/belief exemptions, were statistically significant with all factors, except `polMiss` and `mmrMiss`. Though early ideologies, including protest of the polio vaccine being created from animals may have existed, consistent state regulations have likely made this a nonsignificant issue today. Specifically, polio vaccinations are required by all 50 states and DC for children entering childcare and public schools. Furthermore, religious/belief exemptions allowed students to avoid Measles/Mumps/Rubella vaccinations. However, in recent years, an alarming proportion of outbreaks have encouraged the removal of the exemption. State legislatures in New Jersey, New York, Iowa, Maine, and Vermont have proposed the removal of this exemption. Since polio vaccination is required by all states, this is equally non-significant as being able to avoid mmr vaccination altogether.

Finally, the percentage of student with a medical exemption was statistically significant with all factors except `varMiss`. Similar to the polio vaccination, Varicella (i.e. chicken pox) vaccination is required by all 50 states including DC for all children entering childcare or public schools.

In general, the earlier analysis is a depiction on current societal beliefs, as well as a mirror representation of state laws and guidelines. It will be interesting in recent light, whether state legislatures in New Jersey, New York, Iowa, Maine, and Vermont, will be able to set a precedence among other states. Since ideologies are often re-evaluated - one may ask, should religious belief exempt an infant or child from receiving the best possible health option? Furthermore, religious convictions of one individual or family can jeopardize the health of another. Thus, such discussions and analysis is a fundamental journey to improving the overall health of a society.

Sources

- https://freakonometrics.hypotheses.org/8210
- https://stats.idre.ucla.edu/r/dae/logit-regression/
- https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide.php
- https://www.webmd.com/a-to-z-guides/prevention-15/vaccines/need-hepatitis-vaccines
- https://www.cdc.gov/hepatitis/hav/havfaq.htm
- https://www.cdc.gov/features/hepatitis-b-vaccine/index.html
- https://www.washingtonpost.com/religion/2019/02/21/some-anti-vaccination-parents-cite-religious-exemptions-measles-outbreaks-could-change-that/?utm_term=.d5c3ebef9153
- https://www.cdc.gov/vaccines/vpd/polio/public/index.html
- https://wwwnc.cdc.gov/eid/article/15/6/09-0087_article
- https://www.cdc.gov/vaccines/vpd-vac/varicella/basic-school-daycare-req.htm

Appendix A: histogram signifying overall proportion reporting vaccinations

```
library('ggplot2')
ggplot(
  allSchoolsReportStatus,
  aes(x=reported)
) +
geom_histogram(stat='count')
```

Appendix B: histogram signifying proportion of public and private schools reporting vaccinations.

```
library('ggplot2')
ggplot(
  allSchoolsReportStatus,
  aes(x=reported)
) +
facet_grid(~pubpriv) +
geom_histogram(stat='count')
```

Appendix C: tabulated counts of schools reporting vaccinations.

```
contingency = table(
  allSchoolsReportStatus$reported,
  allSchoolsReportStatus$pubpriv
)

contingency

##
##     PRIVATE PUBLIC
##   N     252    148
##   Y    1397   5584
```

Appendix D: normalized counts of schools reporting vaccinations.

```
prop.table(contingency)

##
##         PRIVATE     PUBLIC
##   N 0.03414172 0.02005148
##   Y 0.18926975 0.75653705
```

Appendix E: time series plots for all vaccination types.

```r
library('reshape2')
df.m = melt(usVaccines,id='X')
ggplot(
  data = df.m, aes(x=X,y=value,colour=variable,group=variable)
) +
  geom_line()
```

Appendix F: decomposed time series plots for all vaccination types.

```r
##
## custom decompose: allows custom title
##
## Note: https://stackoverflow.com/a/43054751
##
plot.decomposed.ts = function(x, title='') {
  xx = x$x
  if (is.null(xx))
    xx = with(x, if (type == 'additive')
      random + trend + seasonal
      else random * trend * seasonal)
  plot(
    cbind(
      observed = xx,
      trend = x$trend,
      seasonal = x$seasonal,
      random = x$random
    ),
    main=title
  )
}

data = subset(usVaccines, select = -c(X))
for(col in names(data)) {
  data.ts = ts(usVaccines[[col]], frequency = 4)
  data.decompose = decompose(data.ts, 'additive')
  plot.decomposed.ts(data.decompose, col)
}
```

Appendix G: high density interval for overall proportions.

```r
library('BEST')
library('BayesFactor')

## Loading required package: coda
```

```
## Loading required package: Matrix

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

schoolReportOut = BESTmcmc(
  allSchoolsReportStatus['reported'][allSchoolsReportStatus$pubpriv=='PUBLIC'
, 1],
  allSchoolsReportStatus['reported'][allSchoolsReportStatus$pubpriv=='PRIVATE
', 1],
  numSavedSteps=1000
)

## Warning in mean.default(y1): argument is not numeric or logical: returning
## NA

## Warning in mean.default(y2): argument is not numeric or logical: returning
## NA

## Warning in var(if (is.vector(x) || is.factor(x)) x else as.double(x), na.r
m = na.rm): Calling var(x) on a factor x is deprecated and will become an err
or.
##   Use something like 'all(duplicated(x)[-1L])' to test for a constant vect
or.

## Warning in var(if (is.vector(x) || is.factor(x)) x else as.double(x), na.r
m = na.rm): Calling var(x) on a factor x is deprecated and will become an err
or.
##   Use something like 'all(duplicated(x)[-1L])' to test for a constant vect
or.

## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 7381
##    Unobserved stochastic nodes: 5
##    Total graph size: 14781
##
## Initializing model
##
##
## Sampling from the posterior distributions:

plot(schoolReportOut)
```

Appendix H:

```
t.test(
  x=reportSample$allvaccs[reportSample$pubpriv == 'PUBLIC'],
  y=reportSample$allvaccs[reportSample$pubpriv == 'PRIVATE']
)

##
##   Welch Two Sample t-test
##
## data:  reportSample$allvaccs[reportSample$pubpriv == "PUBLIC"] and reportS
ample$allvaccs[reportSample$pubpriv == "PRIVATE"]
## t = 1.7274, df = 164.71, p-value = 0.08597
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   -0.3852022  5.7711341
## sample estimates:
## mean of x mean of y
##   89.42484  86.73187
```

Appendix I: trace and density plot for overall vaccination rates.

```
allVacsOut = anovaBF(
  formula = allvaccs ~ pubpriv,
  data=reportSample
)
summary(allVacsOut)

## Bayes factor analysis
## --------------
## [1] pubpriv : 0.8136571 ±0%
##
## Against denominator:
##    Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS

mcmcOut = posterior(allVacsOut, iterations=1000)
summary(mcmcOut)

##
## Iterations = 1:1000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 1000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
```

```
##                       Mean       SD Naive SE Time-series SE
## mu                  88.131  0.6324  0.02000         0.02000
## pubpriv-PRIVATE     -1.305  0.6469  0.02046         0.02046
## pubpriv-PUBLIC       1.305  0.6469  0.02046         0.02046
## sig2               183.587 10.2257  0.32337         0.32337
## g_pubpriv            1.695 19.6505  0.62140         0.62140
##
## 2. Quantiles for each variable:
##
##                       2.5%      25%      50%      75%     97.5%
## mu                 86.90273  87.7052  88.1463  88.5497  89.36568
## pubpriv-PRIVATE    -2.60698  -1.7477  -1.3100  -0.8616  -0.03118
## pubpriv-PUBLIC      0.03118   0.8616   1.3100   1.7477   2.60698
## sig2              164.41069 176.8410 183.1637 190.0335 204.72611
## g_pubpriv           0.03669   0.1019   0.2109   0.5471   4.98085
```

```r
plot(mcmcOut[,'mu'])
```

Appendix J: HDI for overall vaccination rates.

```r
hist(mcmcOut)
abline(v=quantile(mcmcOut, c(0.025)), col='green')
abline(v=quantile(mcmcOut, c(0.975)), col='green')
```

Appendix K:

```r
t.test(
  x=reportSample$medical[reportSample$pubpriv == 'PUBLIC'],
  y=reportSample$medical[reportSample$pubpriv == 'PRIVATE']
)
```

```
##
##   Welch Two Sample t-test
##
## data:  reportSample$medical[reportSample$pubpriv == "PUBLIC"] and reportSa
mple$medical[reportSample$pubpriv == "PRIVATE"]
## t = -1.089, df = 206.02, p-value = 0.2774
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   -0.5457577  0.1573679
## sample estimates:
## mean of x mean of y
## 0.2317120 0.4259069
```

Appendix L: trace and density plot for medical exemptions between public and private schools.

```r
medicalExOut = anovaBF(
  formula = medical ~ pubpriv,
```

```
   data=reportSample
)
summary(medicalExOut)

## Bayes factor analysis
## --------------
## [1] pubpriv : 0.179636 ±0%
##
## Against denominator:
##    Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS

mcmcOut = posterior(medicalExOut, iterations=1000)
summary(mcmcOut)

##
## Iterations = 1:1000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 1000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##                      Mean      SD Naive SE Time-series SE
## mu               0.32930 0.09384 0.002967       0.002967
## pubpriv-PRIVATE  0.09425 0.09497 0.003003       0.003003
## pubpriv-PUBLIC  -0.09425 0.09497 0.003003       0.003003
## sig2             3.78648 0.20079 0.006350       0.006350
## g_pubpriv        0.79433 4.05028 0.128081       0.128081
##
## 2. Quantiles for each variable:
##
##                     2.5%      25%      50%      75%    97.5%
## mu               0.14591  0.26920  0.32718  0.39288 0.51667
## pubpriv-PRIVATE -0.08582  0.02936  0.09205  0.16066 0.27861
## pubpriv-PUBLIC  -0.27861 -0.16066 -0.09205 -0.02936 0.08582
## sig2             3.42830  3.64563  3.77728  3.92133 4.20539
## g_pubpriv        0.03682  0.09174  0.18082  0.45687 4.26969

plot(mcmcOut[,'mu'])
```

Appendix M: HDI for medical exemptions between public and private schools.

```
hist(mcmcOut)
abline(v=quantile(mcmcOut, c(0.025)), col='green')
abline(v=quantile(mcmcOut, c(0.975)), col='green')
```

Appendix N: frequentist t-test for religious/belief exemptions.

```
t.test(
  x=reportSample$religious[reportSample$pubpriv == 'PUBLIC'],
  y=reportSample$religious[reportSample$pubpriv == 'PRIVATE']
)

##
##  Welch Two Sample t-test
##
## data:  reportSample$religious[reportSample$pubpriv == "PUBLIC"] and report
Sample$religious[reportSample$pubpriv == "PRIVATE"]
## t = -1.5412, df = 173.7, p-value = 0.1251
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.0247396  0.3721521
## sample estimates:
## mean of x mean of y
##  3.689237  5.015531
```

Appendix O: bayes factor for religious/belief exemptions.

```
religiousOut = anovaBF(
  formula = religious ~ pubpriv,
  data=reportSample
)
summary(religiousOut)

## Bayes factor analysis
## --------------
## [1] pubpriv : 0.4423991 ±0%
##
## Against denominator:
##   Intercept only
## ---
## Bayes factor type: BFlinearModel, JZS

mcmcOut = posterior(religiousOut, iterations=1000)
summary(mcmcOut)

##
## Iterations = 1:1000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 1000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
```

```
##
##                    Mean      SD Naive SE Time-series SE
## mu                4.3459 0.3788  0.01198         0.01198
## pubpriv-PRIVATE   0.6524 0.3786  0.01197         0.01197
## pubpriv-PUBLIC   -0.6524 0.3786  0.01197         0.01197
## sig2             63.4964 3.5061  0.11087         0.11087
## g_pubpriv         0.6078 1.7955  0.05678         0.08170
##
## 2. Quantiles for each variable:
##
##                      2.5%      25%     50%     75%    97.5%
## mu                3.55382  4.09912  4.3506  4.5951  5.06433
## pubpriv-PRIVATE  -0.07379  0.38779  0.6533  0.8950  1.41969
## pubpriv-PUBLIC   -1.41969 -0.89502 -0.6533 -0.3878  0.07379
## sig2             57.11954 61.02815 63.4182 65.7926 70.86605
## g_pubpriv         0.03716  0.09936  0.1953  0.4488  3.76998
```

```r
plot(mcmcOut[,'mu'])
```

Appendix P: histogram plot using bayes output for religious/belief exemptions.

```r
hist(mcmcOut)
abline(v=quantile(mcmcOut, c(0.025)), col='green')
abline(v=quantile(mcmcOut, c(0.975)), col='green')
```

Appendix Q: logistic regression for overall public and private school prediction.

```r
reportAdjustedSample = reportSample
reportAdjustedSample$pubpriv = as.numeric(reportAdjustedSample$pubpriv) - 1
pubpriv = glm(
  pubpriv ~ conditional + medical + religious,
  data = reportAdjustedSample,
  family = binomial()
)
summary(pubpriv)
```

```
##
## Call:
## glm(formula = pubpriv ~ conditional + medical + religious, family = binomi
al(),
##     data = reportAdjustedSample)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9010   0.5989   0.6148   0.6407   1.2408
##
## Coefficients:
```

```
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.627598   0.126827  12.833   <2e-16 ***
## conditional -0.009361   0.008425  -1.111   0.2665
## medical      -0.035439   0.040256  -0.880   0.3787
## religious    -0.017412   0.010563  -1.648   0.0993 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 671.10  on 697  degrees of freedom
## Residual deviance: 666.58  on 694  degrees of freedom
## AIC: 674.58
##
## Number of Fisher Scoring iterations: 4

plot(pubpriv)
```