Matthew L. Pergolski
IST 772
Dr. Block
12/21/2021

<u>**Final Exam**</u>

<u>**Beginning Statement**</u>
"I produced the material below with no assistance [direct quote from IST 772 class syllabus]." Note:
Homework questions from the book may have been copied/pasted into the document for both the student
and viewer's convenience. Quotes may also have been used from the class textbook.

<u>**Final Exam Question 1**</u>

Question:
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?

Answer/Student Response:

In general, without differencing the data (i.e., applying the diff() command), we can see a general increase
trend in all attributes that associate with the 'usVaccines' data. We can confirm this through the plot()
command as well as the boxplot() command as well. Although the beginning of the time period observes
volatility, all vaccines appear to end at higher values compared to the beginning of the time series. At the
conclusion of the time series, we can see which vaccine(s) have the highest and lowest value by using the
tail() commands. The highest at the end of the interval appears to be 'DTP1' with a rate value of 98 while
'HepB_BD' is by far the lowest with a rate value of just 64. With visually looking at the box plots for
these vaccine rates, we see the 'HepB_BD' also has the 'longer' spread of values, which suggests a more
substantial volatility compared to the other variables.
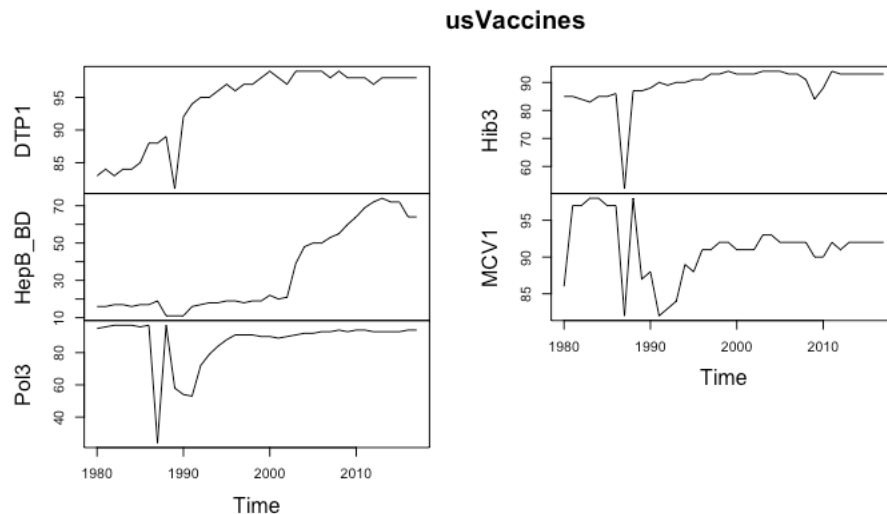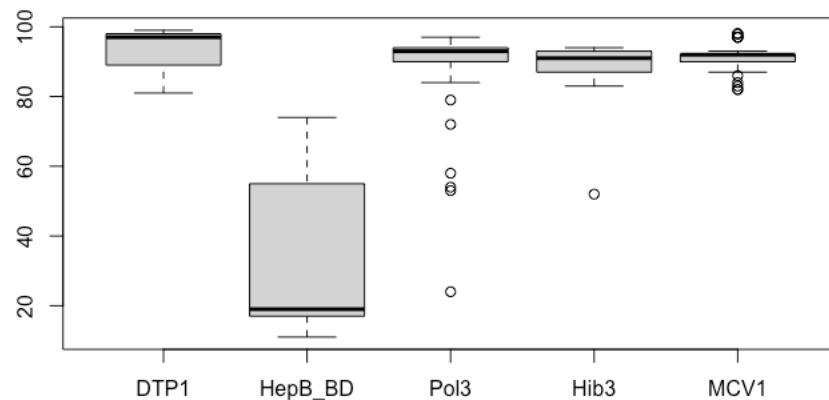
usVaccines

```
>    tail(usVaccines)
     DTP1 HepB_BD Pol3 Hib3 MCV1
[33,]   97      72   93   93   91
[34,]   98      74   93   93   92
[35,]   98      72   93   93   92
[36,]   98      72   93   93   92
[37,]   98      64   94   93   92
[38,]   98      64   94   93   92
```

```
>    summary(usVaccines)
     DTP1           HepB_BD          Pol3            Hib3            MCV1
 Min.   :81.00   Min.   :11.00   Min.   :24.00   Min.   :52.00   Min.   :82.00
 1st Qu.:89.75   1st Qu.:17.00   1st Qu.:90.00   1st Qu.:87.00   1st Qu.:90.00
 Median :97.00   Median :19.00   Median :93.00   Median :91.00   Median :92.00
 Mean   :94.05   Mean   :34.21   Mean   :87.16   Mean   :89.21   Mean   :91.24
 3rd Qu.:98.00   3rd Qu.:54.50   3rd Qu.:94.00   3rd Qu.:93.00   3rd Qu.:92.00
 Max.   :99.00   Max.   :74.00   Max.   :97.00   Max.   :94.00   Max.   :98.00
```

Homework Question 2

Question:
2. 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?

Answer/Student Response:

The following was generated and observed:

```r
# data prep
allSchoolsReportStatus
head(allSchoolsReportStatus)
str(allSchoolsReportStatus)
allSchoolsReportStatus$name <- as.factor(allSchoolsReportStatus$name)
allSchoolsReportStatus$pubpriv <- as.factor(allSchoolsReportStatus$pubpriv)
allSchoolsReportStatus$reported <- as.factor(allSchoolsReportStatus$reported)
str(allSchoolsReportStatus)

# proportion of public schools who reported vaccination data
public.schools <- allSchoolsReportStatus[allSchoolsReportStatus$pubpriv == 'PUBLIC',]
public.schools
str(public.schools)
public.schools.Y <- public.schools[public.schools$reported == 'Y',]
public.schools.Y
public.schools.N <- public.schools[public.schools$reported == 'N',]
public.schools.N

pub.vax.proportion <- length(public.schools.Y$reported) / length(public.schools$reported)
pub.vax.proportion

# 0.97418 (~97%) of public schools reported vaccinations


# proportion of private schools who reported vaccination data
private.schools <- allSchoolsReportStatus[allSchoolsReportStatus$pubpriv == 'PRIVATE',]
private.schools
str(private.schools)
private.schools.Y <- private.schools[private.schools$reported == 'Y',]
private.schools.Y
private.schools.N <- private.schools[private.schools$reported == 'N',]
private.schools.N

length(private.schools.Y$reported)
length(private.schools$reported)

pri.vax.proportion <- length(private.schools.Y$reported) / length(private.schools$reported)
pri.vax.proportion

# 0.8471801 (~85%) of private schools reported vaccinations

# t.test
public.schools$number <- as.numeric( factor(public.schools$reported) ) -1
public.schools$number
private.schools$number <- as.numeric( factor(private.schools$reported) ) -1
private.schools$number

t.test(public.schools$number, private.schools$number)

# bestMCMC
sample(public.schools$number, 100, replace = FALSE)
public.sample <- sample(public.schools$number, 100, replace = FALSE)
public.sample

sample(private.schools$number, 100, replace = FALSE)
pri.sample <- sample(private.schools$number, 100, replace = FALSE)
pri.sample

bayesian.test.problem.2 <- BESTmcmc(public.sample, pri.sample)
bayesian.test.problem.2
summary(bayesian.test.problem.2)
plot(bayesian.test.problem.2)
```

```
>      t.test(public.schools$number, private.schools$number)

       Welch Two Sample t-test

data:  public.schools$number and private.schools$number
t = 13.944, df = 1835.6, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1091376 0.1448623
sample estimates:
mean of x mean of y
0.9741800 0.8471801
```


Difference of Means

mean = 1.66e-07

49.9% < 0 < 50.1%

95% HDI

-6.03e-05   6.05e-05

$\mu_1 - \mu_2$
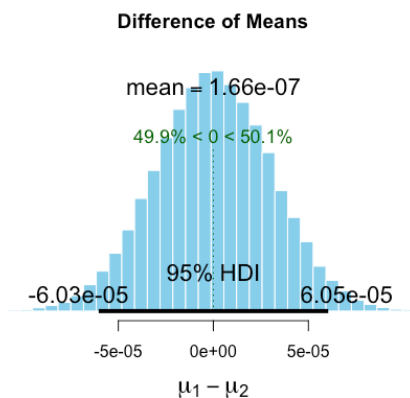
0.97418 (~97%) of public schools reported vaccinations. 0.8471801 (~85%) of private schools reported vaccinations. In order to generate this statistic, the 'allSchoolsReportStatus' data was split into two groups: public and private. From here, they were split into two more groups – those who reported vaccinations and those who did not. The number of schools (for both public and private, respectively) that reported vaccinations were divided by the total number of schools (public or private, respectively). Please reference the R-code script for additional information.

The observed p-value from the frequentist t-test provides a value of 2.2e-16, which suggests statistical significance. The confidence interval also ranges from 0.1091376 to 0.1448623. Although this frequentist confidence interval is a 'long-term' solution (which may or may not apply to this specific time the test was ran), we do not see that this interval spans across zero, which also gives us confidence that we can reject the null hypothesis. There does seem to be a credible difference in overall reporting between public and private schools. In order to run the BESTmcmc() command the dataset was trimmed and sampled. From the results of the Bayesian test, we see a mean difference that does unfortunately span across zero, with the 95% HDI showing ~-6 on the lower bound and ~6 on the upper bound, as shown above.

I believe the trimmed dataset unfortunately may have skewed the results, so the frequentist perspective will be viewed as more legitimate, with statistical significance over the smaller mean difference. Regardless, the difference does not seem to be very large.

Homework Question 3

Question:

3. 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)?

Answer/Student Response:

The following code was generated and observed:

```
str(districts)
summary(districts)
head(districts)
plot(districts)

str(allSchoolsReportStatus)
View(allSchoolsReportStatus)

allSchoolsReportStatus[allSchoolsReportStatus$name == 'Cajon Valley Union']

districts[districts$DistrictName == 'Cajon Valley Union',]


colnames(allSchoolsReportStatus)
colnames(districts)
districts.2013 <- districts
districts.2013$name <- toupper(districts[,1])
districts.2013$name

districts.2013 <- merge(allSchoolsReportStatus, districts.2013, by = 'name')
districts.2013

districts.2013.public <- districts.2013[districts.2013$pubpriv == 'PUBLIC',]
districts.2013.public

plot(districts.2013.public$PctUpToDate)

100 - mean(districts.2013.public$WithoutDTP)
# 91.56842% for DTP in 2013 public schools
100 - mean(districts.2013.public$WithoutPolio)
# 91.85614for Polio in 2013 public schools
100 - mean(districts.2013.public$WithoutMMR)
# 91.37544% for MMR in 2013 public schools
100 - mean(districts.2013.public$WithoutHepB)
# 93.13684 for HepB in 2013 public schools
```

```
>     mean(districts.2013.public$WithoutDTP)
[1] 8.431579
>     100 - mean(districts.2013.public$WithoutDTP)
[1] 91.56842
>     100 - mean(districts.2013.public$WithoutPolio)
[1] 91.85614
>     100 - mean(districts.2013.public$WithoutMMR)
[1] 91.37544
>     100 - mean(districts.2013.public$WithoutHepB)
[1] 93.13684
```

A merge took place of the allSchoolsReportStatus and the districts dataset. This merge took place in order to identify vaccination rates for 2013 public school districts. From here, since the rates are categorized as a percentage of those who did not receive the vaccines, the rates were ultimately subtracted from 100 to provide a rate of those who did get vaccinated. We see that the vaccination rate, on average, was 91.56842% for DTP in 2013 public schools, 91.37544% for MMR in 2013 public schools, 91.37544% for MMR in 2013 public schools, and 93.13684 for HepB in 2013 public schools. These rates are comparableto the ending of the dataset except for HepB_BD -- where the average of HepB is significantly greater than the last remaining value in the usVaccines dataset for HepB.

Homework Question 4

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?

Answer/Student Response:

```
>       cor(districts.vaxes)
            WithoutDTP WithoutPolio WithoutMMR WithoutHepB
WithoutDTP   1.0000000    0.9803893  0.9760780    0.8858368
WithoutPolio 0.9803893    1.0000000  0.9639514    0.9020601
WithoutMMR   0.9760780    0.9639514  1.0000000    0.8872849
WithoutHepB  0.8858368    0.9020601  0.8872849    1.0000000
```

```
>       cor.test(districts.vaxes$WithoutDTP, districts.vaxes$WithoutPolio)

        Pearson's product-moment correlation

data:  districts.vaxes$WithoutDTP and districts.vaxes$WithoutPolio
t = 131.43, df = 698, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9772859 0.9830723
sample estimates:
      cor
0.9803893
```

```
>      bfCorTest(districts.vaxes$WithoutDTP, districts.vaxes$WithoutPolio)
  |=========================================================================================| 100%
0                                                                                            %
|----|----|----|----|----|----|----|----|----|----|
*************************************************|

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
     9.804e-01      7.431e-03      7.431e-05      7.431e-05

2. Quantiles for each variable:

  2.5%    25%    50%    75%  97.5%
0.9659 0.9753 0.9804 0.9854 0.9949

Bayes factor analysis
--------------
[1] rhoNot0 : 4.937621e+489 ±0.01%

Against denominator:
  Intercept only
---
Bayes factor type: BFlinearModel, JZS
```

Among the districts, when looking at a correlation matrix, we see high degrees of correlation between the different attributes variables. Since these variables are categorized as those who did not receive the vaccine, this suggests that, if one student is to miss the DTP vaccine, for example, it is highly likely they will also miss the Polio vaccine (correlation value of 0.9803893). If we dive deeper into this example of DTP and Polio, we find that both the frequentist approach as well as Bayesian point to similar results (that the variables are highly correlated). We see a p-value for the frequentist approach come out to be 2.2e-16, which suggests statistical significance when comparing it to our conventional threshold of 0.05. When performing the Bayesian test, we see a rhoNot0 value of 4.937621e+489 to 1, which is extremely high. If we remember our standard for determining strong odds for this test, we would indicate anything greater than 150 to 1 would be strong odds. To reiterate, although we did just two vaccines in this example, we can see other correlation measures in the standard cor() command.

Homework Question 5

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

Answer/Student Response:

The following as generated and observed:

```
districts.lm.data <- districts[,2:13]

linear.model.5 <- lm(DistrictComplete ~ ., districts.lm.data)
summary(linear.model.5)

districts.lm.data$number <- as.numeric( factor(districts.lm.data$DistrictComplete) ) -1
lmBF(number ~ ., districts.lm.data)
```

```
>    summary(linear.model.5)

Call:
lm(formula = DistrictComplete ~ ., data = districts.lm.data)

Residuals:
    Min      1Q   Median      3Q      Max
-1.01644  0.01213  0.04422  0.07701  0.67410

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      8.037e-01  2.916e-01   2.757    0.006 **
WithoutDTP      -1.757e-03  5.255e-03  -0.334    0.738
WithoutPolio    -4.601e-03  4.413e-03  -1.043    0.297
WithoutMMR       4.637e-03  4.379e-03   1.059    0.290
WithoutHepB      2.158e-03  3.104e-03   0.695    0.487
PctUpToDate      2.215e-03  2.942e-03   0.753    0.452
PctBeliefExempt  1.408e-03  2.645e-03   0.532    0.595
PctChildPoverty  1.660e-03  1.482e-03   1.120    0.263
PctFreeMeal     -8.751e-04  5.495e-04  -1.592    0.112
PctFamilyPoverty -3.318e-03  2.082e-03  -1.594    0.111
Enrolled         2.481e-04  3.759e-05   6.599 8.24e-11 ***
TotalSchools    -2.492e-02  3.485e-03  -7.152 2.20e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2238 on 688 degrees of freedom
Multiple R-squared:  0.127,     Adjusted R-squared:  0.113
F-statistic: 9.095 on 11 and 688 DF,  p-value: 2.722e-15
```

When comparing all variables to the 'DistrictComplete' attribute, we see that 2 variables have statistical significance and may be sufficient predictors: (1) 'Enrolled' (p-value = 8.24e-11) and (2) 'TotalSchools (p-value = 2.20e-12).' The intercept is also significant with a p-value at 0.006. Although the overall p-value shown is 2.2e-16, we see that the R and R-squared values are around 0.1, which does not suggest a strong model.

Homework Question 6

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

Answer/Student Response:

The following was generated and observed:

```
colnames(districts.lm.data)

linear.model.6 <- lm(PctUpToDate ~ ., districts.lm.data)
summary(linear.model.6)

bayesian.model.6 <- lmBF(PctUpToDate ~ WithoutDTP + WithoutMMR + PctBeliefExempt, data = districts.lm.data,
                    posterior = TRUE, iterations = 10000)
summary(bayesian.model.6)
```

```
>      summary(linear.model.6)

Call:
lm(formula = PctUpToDate ~ ., data = districts.lm.data)

Residuals:
    Min      1Q  Median      3Q     Max
-42.697  -0.342   0.460   1.032  12.256

Coefficients: (1 not defined because of singularities)
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        9.840e+01  5.903e-01 166.700  < 2e-16 ***
WithoutDTP        -5.004e-01  6.534e-02  -7.658 6.41e-14 ***
WithoutPolio       7.926e-02  5.712e-02   1.388    0.166
WithoutMMR        -7.680e-01  4.863e-02 -15.791  < 2e-16 ***
WithoutHepB       -5.686e-02  4.017e-02  -1.416    0.157
DistrictCompleteTRUE  3.717e-01  4.936e-01   0.753    0.452
PctBeliefExempt    2.245e-01  3.319e-02   6.765 2.86e-11 ***
PctChildPoverty    1.838e-02  1.920e-02   0.957    0.339
PctFreeMeal        6.691e-03  7.126e-03   0.939    0.348
PctFamilyPoverty  -2.312e-02  2.700e-02  -0.856    0.392
Enrolled           1.101e-05  5.022e-04   0.022    0.983
TotalSchools      -3.065e-03  4.679e-02  -0.066    0.948
number                    NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.899 on 688 degrees of freedom
Multiple R-squared:  0.9452,    Adjusted R-squared:  0.9444
F-statistic:  1079 on 11 and 688 DF,  p-value: < 2.2e-16

>
```

```
>      summary(bayesian.model.6)

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              88.0794  0.10954 0.0010954      0.0010954
WithoutDTP      -0.4425  0.04997 0.0004997      0.0004997
WithoutMMR      -0.7795  0.04761 0.0004761      0.0004761
PctBeliefExempt  0.1967  0.02151 0.0002151      0.0002151
sig2             8.4460  0.45810 0.0045810      0.0045647
g                8.4205 13.34130 0.1334130      0.1334130

2. Quantiles for each variable:

                   2.5%     25%     50%     75%   97.5%
mu              87.8661 88.0068 88.0803 88.1533 88.2923
WithoutDTP      -0.5387 -0.4756 -0.4422 -0.4091 -0.3473
WithoutMMR      -0.8691 -0.8108 -0.7798 -0.7487 -0.6883
PctBeliefExempt  0.1540  0.1824  0.1968  0.2113  0.2383
sig2             7.6082  8.1344  8.4311  8.7374  9.3698
g                1.5591  3.1582  5.0308  8.7348 36.6174

>
```

When comparing all variables to the 'PctUpToDate' attribute, we see that 3 variables have statistical significance and may be sufficient predictors: (1) 'WithoutDTP' (p-value = 6.41e-14) and (2) 'WithoutMMR (p-value = WithoutMMR), (3) 'PctBeliefExempt' (p = 2.86e-11). The intercept is also significant with a p-value at 2e-16. For the the R and R-squared values, we see values of are around 0.94, which suggests a relatively strong model. When running the Bayesian version of this test on this predicting variables, we see similar results and do not see the HDI span across zero for any of the chosen attributes in the summary(bayesian.model.6) function.

Homework Question 7

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

Answer/Student Response:

The following was generated and observed:

```
colnames(districts.lm.data)

linear.model.7 <- lm(PctBeliefExempt ~ ., districts.lm.data)
summary(linear.model.7)

bayesian.model.7 <- lmBF(PctBeliefExempt ~ WithoutDTP + WithoutHepB + PctUpToDate + PctFreeMeal,
                         data = districts.lm.data, posterior = TRUE, iterations = 10000)
summary(bayesian.model.7)
```

```
>       summary(linear.model.7)

Call:
lm(formula = PctBeliefExempt ~ ., data = districts.lm.data)

Residuals:
    Min      1Q  Median      3Q     Max
-41.433  -0.550   0.236   0.881  13.103

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          -2.791e+01  4.088e+00  -6.827 1.91e-11 ***
WithoutDTP            2.452e-01  7.515e-02   3.263  0.00116 **
WithoutPolio         1.138e-01  6.349e-02   1.793  0.07348 .
WithoutMMR          -8.160e-02  6.308e-02  -1.294  0.19622
WithoutHepB          8.242e-01  3.186e-02  25.865  < 2e-16 ***
PctUpToDate          2.778e-01  4.107e-02   6.765 2.86e-11 ***
DistrictCompleteTRUE 2.924e-01  5.493e-01   0.532  0.59472
PctChildPoverty      4.065e-02  2.132e-02   1.907  0.05700 .
PctFreeMeal         -2.276e-02  7.885e-03  -2.886  0.00402 **
PctFamilyPoverty    -3.657e-03  3.005e-02  -0.122  0.90317
Enrolled             3.419e-04  5.585e-04   0.612  0.54061
TotalSchools        -3.555e-02  5.203e-02  -0.683  0.49469
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.225 on 688 degrees of freedom
Multiple R-squared:  0.8512,    Adjusted R-squared:  0.8488
F-statistic: 357.9 on 11 and 688 DF,  p-value: < 2.2e-16

> |
```

```
>       summary(bayesian.model.7)

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            5.498082 0.123717 1.237e-03       1.232e-03
WithoutDTP    0.295673 0.046554 4.655e-04       4.800e-04
WithoutHepB   0.841539 0.029108 2.911e-04       2.940e-04
PctUpToDate   0.311535 0.034437 3.444e-04       3.444e-04
PctFreeMeal  -0.009159 0.005277 5.277e-05       5.277e-05
sig2         10.538300 0.565199 5.652e-03       5.757e-03
g             1.880249 2.207313 2.207e-02       2.371e-02

2. Quantiles for each variable:

                  2.5%       25%       50%       75%      97.5%
mu             5.25805   5.41336   5.500041   5.581114   5.739505
WithoutDTP     0.20387   0.26448   0.295267   0.326825   0.385818
WithoutHepB    0.78528   0.82199   0.841631   0.861318   0.899552
PctUpToDate    0.24408   0.28812   0.311314   0.334929   0.379078
PctFreeMeal   -0.01948  -0.01269  -0.009132  -0.005584   0.001276
sig2           9.48139  10.15247  10.523061  10.897878  11.706076
g              0.43292   0.84489   1.304959   2.109276   6.824780

>
```

When comparing all variables to the 'PctBeliefExempt' attribute, we see that 4 variables have statistical significance and may be sufficient predictors: (1) 'WithoutDTP' (p-value = 0.00116), (2) 'WithoutHepB' (p-value = 2e-16), (3) 'PctUpToDate' (p = 2.86e-11), and (4) 'PctFreeMeal' (p = 0.00402). The intercept is also significant with a p-value at 1.91e-11. For the R and R-squared figures, we see values of 0.84/0.85, which suggests a relatively strong model. When running the Bayesian version of this test on this predicting variables, we see similar results and most of the variables do not have an HDI that span across zero in the summary(bayesian.model.7) function. It is important to note that we do see the 'PctFreeMeal' variable have an HDI of -0.01943 on the lower bound and 0.001079 on the upper bound, which unfortunately spans across zero – this gives us less confidence that this is a true predictor for the 'PctBeliefExempt' attribute.

Homework Question 8

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?

Answer/Student Response:

All in all, we see that vaccine rates are generally increasing over time, which is positive news for the public as a whole in addition to the legislator's office. Although most vaccines have rates in the 90s,

there is one, 'HepB' that unfortunately is around the rate of ~ 60. The legislator should strive to continue these upward trends and achieve a 100% vaccination rate for all vaccines.

A place to focus on in terms of overall vaccine rate is increasing private schools who report their aggregate vaccination status. Although ~97% of public schools reported vaccinations, it was discovered that only ~85% of private schools reported their rates. If we gain additional information in these reporting figures, we may find that vaccination rates are higher overall. An overall statistical significance was found in terms of the difference in reporting data for public versus private schools, with a p-value equating to 2.2e-16. To reiterate, we should attempt to collect and report as much data as we can – the more data, the better and more concrete our overall results will be.

Moreover, when taking into consideration the analysis that was done on the 'districts' dataset, we generally found high correlation between the various vaccines that were represented in the data, and found that some may also be good predictors. To share a few examples from the above analysis, we observed statistical significance when running our linear models and identifying 'PctBeliefExempt' versus the 'WithoutHepB' variable – as well as the 'PctBeliefExempt' ~ 'PctFreeMeal variables. Generally when running Bayesian techniques on these tests, we also found most had a 95% HDI that did not span zero, which only strengthened our beliefs about the validity of our results. When conducting a Bayesian (bfCorTest) test, the correlated results were found to be 4.937621e+489:1, which is exceptional odds by our standard (i.e., results greater than 150:1 is considered 'strong' odds). As a reminder, this customer function was introduced in the IST 772 class textbook that allowed us to find an Bayesian developed odds figure that would help place results in perspective in terms of statistical significance. The full function is defined as follows:

```
bfCorTest <- function (x,y) #Get r from BayesFactor
  {
    zx <- scale(x) #Standardize X
    zy <- scale(y) #Standardize Y
    zData <- data.frame(x=zx,rhoNot0=zy) #Put in a data frame
    bfOut <- generalTestBF(x ~ rhoNot0, data=zData) #linear coefficient
    mcmcOut <- posterior(bfOut,iterations=10000) #posterior samples
    print(summary(mcmcOut[,"rhoNot0"])) #Show the HDI for r
    return(bfOut) #Return Bayes factor object
```

I believe enough information was discovered that allows us to believe that, generally, if a student misses one vaccine, they are more likely to miss other vaccines as well (even if only a few others). Since there appeared to be statistical evidence that suggested those children who qualified for free meals are likely to miss (some) vaccines, we may be able to infer that this socio-economic group may be struggling to prioritize health vaccinations due to financial limitations.

Because of the above, financial assistance should be provided to these families so they can begin to prioritize vaccinations within their children. To be brief, educational programs and other forms of information (with an emphasis on financial assistance) may help increase vaccination rates for those belonging to this specific group. Similar to COVID-19 vaccines now, more should be allocated to ensure vaccinations in children increase (while taking the above information into account).

In terms of which variables predict whether or not a district's reporting was complete, the following were found to be statistically significant: 'Enrolled' and 'TotalSchools'. Subsequently, 'WithoutDTP,' 'WithoutMMR,' and 'PctBeliefExempt' were found to predict 'PctUpToDate.' Lastly, 'WithoutDTP,' 'WithoutHepB,' 'PctUpToDate,' and 'PctFreeMeal' were all found to be statistically significant in predicting 'PctBeliefExempt.'

As more data becomes available, we hope to learn more insights on how vaccination data within the 'districts' dataset is related.