###

#IST772, Final Exam Heading

Student name: David Primrose

Final Exam Week: 11

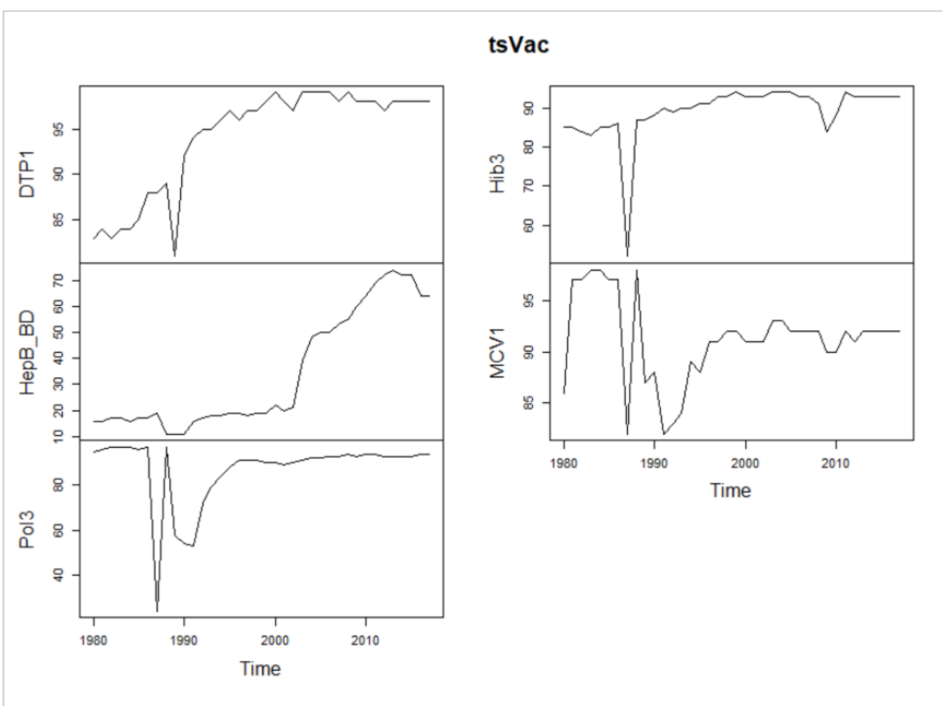
Date due: September 16, 2020

Attribution statement: I did this Final work by myself, with help from the book and the professor

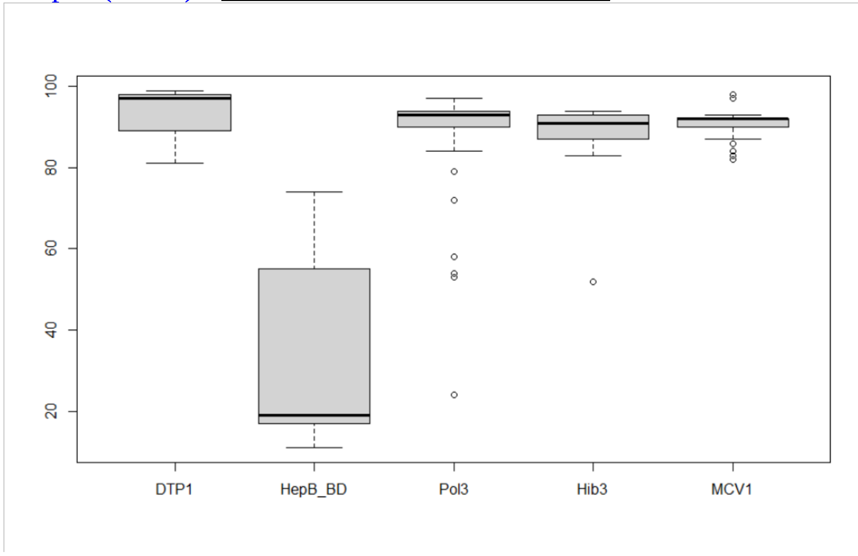
Problem 1:

1. Vaccination rates, except for MCV1 have trended higher year over year higher with some variability included. Most of the major variability occurred after 1985 then subsequently trending higher. In the case of MCV1 it trended higher before leveling off. The greatest volume of vaccines and highest rate is DTP at 3574 vaccines and maximum rate of 99 and a median of a high 97. The least in volume and compliance is HepB_BD at 1300 vaccines, a maximum rate of 74 but a median of only 19. It is unclear whether reason for this is the this is often done at birth and not performed again unless a student has been exposed or is suspected of having hepatitis. Some poor students may not have had access to this test at birth and it may not be priority now leading to a lower level of compliance. The greatest volatility is the Pol3 vaccine with starting count of 24 and a max count of 97., The highest rate of compliance among the vaccines is

plot.ts(tsVac)



boxplot(tsVac) - Of Mean Vaccine Rates in US



Greatest Volatility"

```
> max(tsVac[ ,'DTP1']) - min(tsVac[ ,'DTP1'])
[1] 18
> max(tsVac[ ,"HepB_BD"]) - min(tsVac[ ,"HepB_BD"])
[1] 63
> max(tsVac[ ,'Pol3']) - min(tsVac[ ,'Pol3'])
[1] 73
> max(tsVac[ ,'Hib3']) - min(tsVac[ ,'Hib3'])
[1] 42
> max(tsVac[ ,'MCV1']) - min(tsVac[ ,'MCV1DTP1'])
```

SUM

```
> sum(tsVac[ ,'DTP1'])
[1] 3574
> sum(tsVac[ ,"HepB_BD"])
[1] 1300
> sum(tsVac[ ,'Pol3'])
[1] 3312
> sum(tsVac[ ,'Hib3'])
[1] 3390
> sum(tsVac[ ,'MCV1'])
[1] 3467
```

```
> summary(tsVac)
      DTP1          HepB_BD          Pol3           Hib3
 Min.   :81.00   Min.   :11.00   Min.   :24.00   Min.   :52.00
 1st Qu.:89.75   1st Qu.:17.00   1st Qu.:90.00   1st Qu.:87.00
 Median :97.00   Median :19.00   Median :93.00   Median :91.00
```

```
 Mean   :94.05   Mean   :34.21   Mean   :87.16   Mean   :89.21  
 3rd Qu.:98.00   3rd Qu.:54.50   3rd Qu.:94.00   3rd Qu.:93.00  
 Max.   :99.00   Max.   :74.00   Max.   :97.00   Max.   :94.00  
      MCV1      
 Min.   :82.00  
 1st Qu.:90.00  
 Median :92.00  
 Mean   :91.24  
 3rd Qu.:92.00  
 Max.   :98.00  
```

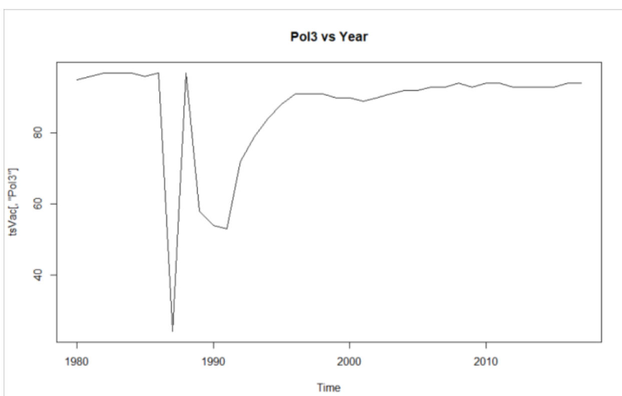
```
> sum(totalVac)
[1] 15043
```

Correlation

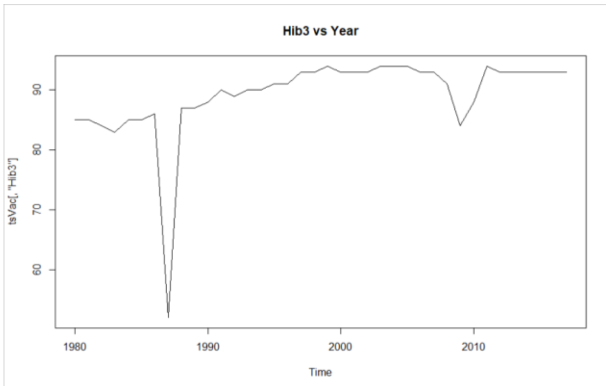
```
> cor(tsVac)
              DTP1   HepB_BD      Pol3      Hib3       MCV1
DTP1     1.0000000 0.5905157 0.1730162 0.5593254 -0.2095028
HepB_BD  0.5905157 1.0000000 0.3255994 0.3282156  0.0819240
Pol3     0.1730162 0.3255994 1.0000000 0.6025809  0.7308557
Hib3     0.5593254 0.3282156 0.6025809 1.0000000  0.2108128
MCV1    -0.2095028 0.0819240 0.7308557 0.2108128  1.0000000
```

Individual Vaccine trend lines:

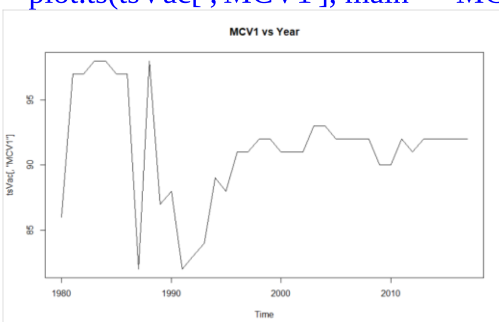
```
> plot.ts(tsVac[ ,'Pol3'], main = "Pol3 vs Year")
```



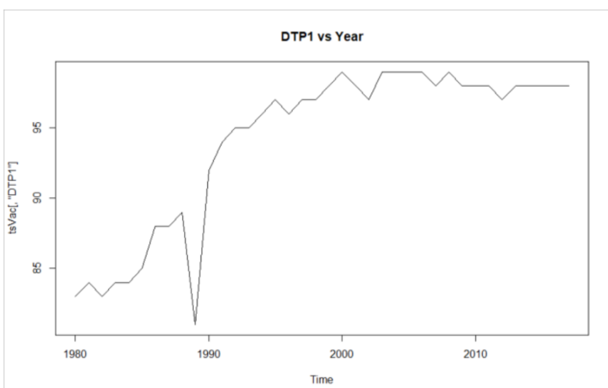

```
> plot.ts(tsVac[ ,'Hib3'], main = "Hib3 vs Year")
```



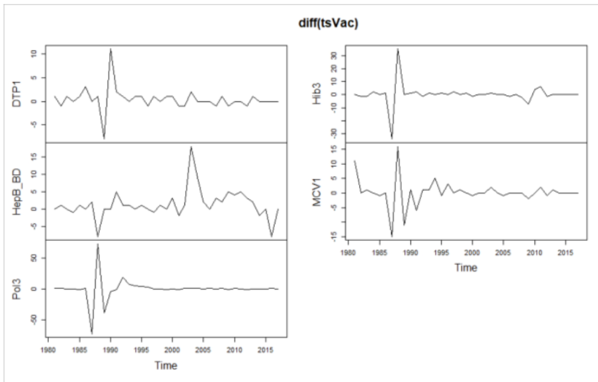
> plot.ts(tsVac[,'MCV1'], main = "MCV1 vs Year")



> plot.ts(tsVac[,'DTP1'], main = "DTP1 vs Year")



plot(diff(tsVac))



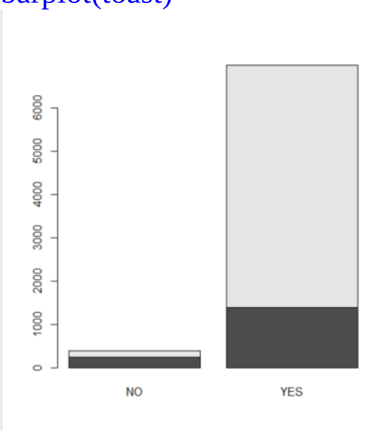

```
> ##################################################
> anovaVac <- aov(DTP1 ~ HepB_BD + Pol3 + Hib3 + MCV1, data = tsVac1)
> summary(anovaVac)
            Df Sum Sq Mean Sq F value   Pr(>F)
HepB_BD      1  444.2   444.2  29.912 4.62e-06 ***
Pol3         1    0.5     0.5   0.036   0.8515
Hib3         1  292.0   292.0  19.662 9.67e-05 ***
MCV1         1   47.1    47.1   3.170   0.0842 .
Residuals   33  490.1    14.9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Problem 2 – Portion of Public Schools reporting (using the "Toast" matrix with margins)

The proportion of public schools reporting student vaccination data is 76% with 19% of private schools reporting as well from a of total 7381 possible. This is a large variance in schools vaccine reporting by type of school. Additionally the percentage of private schools reporting amongst all private schools is 85% vs 97% of public schools amongst all public schools. This is a reporting representing a gap of 12% which is also significant. The overall total of reporting by combined schools is 95% reporting vs 5% not reporting.

```
> toast <- matrix(c(252,1397,148,5584),ncol=2,byrow=TRUE)
# Create a two-column structure using the matrix() command
> colnames(toast) <- c("NO","YES") # Label the columns
> rownames(toast) <- c("PRIVATE","PUBLIC") # Label the rows
> toast <- as.table(toast) # Convert from metric to table
> toast # Show the table on the console

         NO  YES
PRIVATE  252 1397
```

```
PUBLIC   148 5584
> margin.table(toast) # This is the grand total of vaccines
[1] 7381
> margin.table(toast,1) # These are the marginal totals for rows between school types
PRIVATE  PUBLIC
   1649    5732
> margin.table(toast,2) # These are the marginal totals for NO/YES decision totals
  NO  YES
 400 6981
> toastProbs <- toast/margin.table(toast) # Calculate vaccine probabilities
> toastProbs # Report probabilities to console
                NO         YES
PRIVATE 0.03414172  0.18926975
PUBLIC  0.02005148  0.75653705


> USMar <- margin.table(toast,2)/margin.table(toast)
> USMar
       NO       YES
0.0541932 0.9458068

barplot(toast)
```



Problem 3 – Vaccine Rates

The vaccine rates for California individual vaccines in 2013 are as follows:

```
> DTP 98.3
[1] 0.9827633
> Polio 98.3
[1] 0.9834804
> MMR 98.3
[1] 0.9828056
> HepB 98.7
[1] 0.9869602
```

Comparison between California and US: The State of California requires all children attending public or private school to receive the doctor-recommended immunizations for vaccine-preventable diseases. These begin at birth and carries through to age six. There are 10 vaccines for babies and children for 14 diseases. The US vaccine rates are comparable California but with a 5-10% difference of which 6% is split between those with an exemption and those who have slipped by without vaccination.

Vaccine	California	US
DTP	98.3	94.05
Polio	98.3	87.16
MMR	98.3	91.24
HepB	98.7	34.21

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

California Totals

```
> DTP <- 1-(sum(casch$WithoutDTP)/sum(casch$Enrolled))
> Polio <- 1-(sum(casch$WithoutPolio)/sum(casch$Enrolled))
> MMR <- 1-(sum(casch$WithoutMMR)/sum(casch$Enrolled))
> HepB <-1-(sum(casch$WithoutHepB)/sum(casch$Enrolled))
> DTP
[1] 0.9827633
> Polio
[1] 0.9834804
> MMR
[1] 0.9828056
```

```
> HepB
[1] 0.9869602

> sum(casch$PctUpToDate)
[1] 61281
> sum(casch$Enrolled)
[1] 425313
> sum(casch$WithoutDTP)
[1] 7331
> sum(casch$WithoutPolio)
[1] 7026
> sum(casch$WithoutMMR)
[1] 7313
> sum(casch$WithoutHepB)
[1] 5546
> sum(casch$WithoutDTP)
[1] 7331
```

Problem 4 – Vaccine relationships

The vaccine relationships for California individual vaccines in 2013 are as follows:

The highest correlation is not a DTP vaccine is highly correlated to also not having a Polio vaccine at 98%. The next highest is DTP and MMR. If you don't have DTP, you are 98% likely to also not have had an MMR vaccine. This is followed by MMR and Polio at 97%. But across all schools the unvaccinated rates are still only 6%. With Anova the Sum of Squares for variance shows very small variability between groups but a large 34.14 variance within groups with 688 remaining degrees of freedom. The WithoutHepB is the only vaccine that has a significant result.

```
> cor(casch1)
                 WithoutDTP WithoutPolio  WithoutMMR WithoutHepB PctUpToDate PctBeliefExempt PctChildPoverty
WithoutDTP       1.00000000   0.98145070  0.97689096  0.88828720 -0.95797124      0.79743843     -0.20286052
WithoutPolio     0.98145070   1.00000000  0.96494816  0.90346998 -0.93790566      0.81905792     -0.20436907
WithoutMMR       0.97689096   0.96494816  1.00000000  0.89019495 -0.96605673      0.78573694     -0.20219944
WithoutHepB      0.88828720   0.90346998  0.89019495  1.00000000 -0.84015725      0.91982413     -0.21928009
PctUpToDate     -0.95797124  -0.93790566 -0.96605673 -0.84015725  1.00000000     -0.71726000      0.20839878
PctBeliefExempt  0.79743843   0.81905792  0.78573694  0.91982413 -0.71726000      1.00000000     -0.18315310
PctChildPoverty -0.20286052  -0.20436907 -0.20219944 -0.21928009  0.20839878     -0.18315310      1.00000000
PctFreeMeal     -0.25017036  -0.26005871 -0.25384625 -0.30729342  0.24788803     -0.29361083      0.74080749
PctFamilyPoverty -0.24835928 -0.25185271 -0.24509303 -0.27096427  0.24435979     -0.23444041      0.84867209
Enrolled        -0.06756962  -0.06949552 -0.07149217 -0.07796188  0.05859106     -0.08825678      0.03049929
TotalSchools    -0.05536972  -0.05747251 -0.06012823 -0.06695351  0.04671335     -0.08037692      0.02619772
                PctFreeMeal PctFamilyPoverty    Enrolled TotalSchools
WithoutDTP      -0.25017036      -0.24835928 -0.06756962  -0.05536972
WithoutPolio    -0.26005871      -0.25185271 -0.06949552  -0.05747251
WithoutMMR      -0.25384625      -0.24509303 -0.07149217  -0.06012823
WithoutHepB     -0.30729342      -0.27096427 -0.07796188  -0.06695351
PctUpToDate      0.24788803       0.24435979  0.05859106   0.04671335
PctBeliefExempt -0.29361083      -0.23444041 -0.08825678  -0.08037692
PctChildPoverty  0.74080749       0.84867209  0.03049929   0.02619772
PctFreeMeal      1.00000000       0.71374092  0.05997848   0.05588002
PctFamilyPoverty 0.71374092       1.00000000  0.04901523   0.04288837
Enrolled         0.05997848       0.04901523  1.00000000   0.99467938
TotalSchools     0.05588002       0.04288837  0.99467938   1.00000000
>
```

```
> regOutz <- aov(D_Comp ~ WithoutDTP + WithoutPolio + WithoutMMR + WithoutHepB + PctUpToDate + PctBeliefExempt + PctChildPoverty + PctFreeMeal + PctFamilyPoverty + Enrolled + TotalSchools, data=casch)
> regOutz
Call:
  aov(formula = D_Comp ~ WithoutDTP + WithoutPolio + WithoutMMR +  WithoutHepB + PctUpToDate + PctBeliefExempt + PctChildPoverty +  PctFreeMeal + PctFamilyPoverty + Enrolled + TotalSchools,
  data = casch)

Terms:
W/O_DTP W/O_Polio W/O_MMR W/O_HepB %UpToDate %BelExpt %ChildPov
Sum 0.15161 0.00002  0.09688  0.21791  0.02151  0.02686  .16968
DF   1       1         1       1         1       1        1
       %FreeMeal %FamilyPov Enrolled TotalSchools Residuals
Sum       0.17858    0.11693  1.48245   1.99448     34.14168
DFm          1          1        1        1          688

Residual standard error: 0.2227657
Estimated effects may be unbalanced


> summary(regOutz)
              Df Sum Sq Mean Sq F value   Pr(>F)
WithoutDTP        1   0.15  0.1516   3.055   0.0809 .
WithoutPolio      1   0.00  0.0000   0.000   0.9861
WithoutMMR        1   0.10  0.0969   1.952   0.1628
WithoutHepB       1   0.22  0.2179   4.391   0.0365 *
PctUpToDate       1   0.02  0.0215   0.433   0.5106
```

```
PctBeliefExempt    1   0.03  0.0269   0.541   0.4622
PctChildPoverty    1   0.17  0.1697   3.419   0.0649 .
PctFreeMeal        1   0.18  0.1786   3.599   0.0582 .
PctFamilyPoverty   1   0.12  0.1169   2.356   0.1252
Enrolled           1   1.48  1.4824  29.873 6.46e-08 ***
TotalSchools       1   1.99  1.9945  40.191 4.17e-10 ***
Residuals        688  34.14  0.0496
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Prescriptive Analysis

Problem 5:

Whether or not a child was enrolled, and part of the total school population were the most significant predictors of a district being complete in their vaccinations. This is due to the California requirement that all children be vaccinated.

https://www.cde.ca.gov/ls/he/hn/cefimmunization.asp

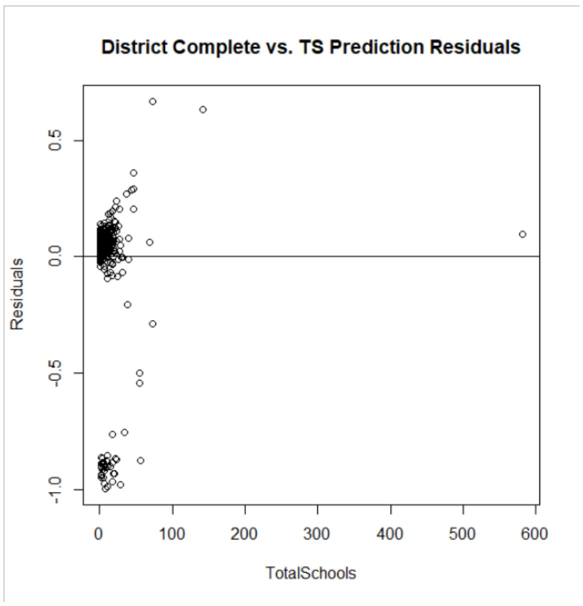
If a child can make it into the California school system, the likelihood of them being complete on their vaccinations raises is significant.

On a national level the non-vaccinated https://www.cdc.gov/nchs/fastats/immunize.htm represent 5% if the student population. Bayesian and Logistic confirmations follow below:

```
> regOutp1

Call:
lm(formula = D_Comp ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty + Enrolled + Tota
lSchools, data = casch1)

Coefficients:
(Intercept)  %ChildPov  %FreeMeal  %FamilyPov  Enrolled  Tot_Sch
0.9990600   0.0014958 0.0006982  -0.0028331  .0002238 -0.0226560

> summary(regOutp1)

Call:
lm(formula = D_Comp ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty + Enrolled + Tota
lSchools, data = casch1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.99574  0.02183  0.04303  0.06875  0.66803
```

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      9.991e-01  1.993e-02  50.117  < 2e-16 ***
PctChildPoverty   1.496e-03  1.445e-03   1.035    0.301
PctFreeMeal      -6.982e-04  5.275e-04  -1.324    0.186
PctFamilyPoverty -2.833e-03  2.071e-03  -1.368    0.172
Enrolled          2.238e-04  3.698e-05   6.053 2.32e-09 ***
TotalSchools     -2.266e-02  3.409e-03  -6.647 6.06e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2231 on 694 degrees of freedom
Multiple R-squared:  0.1049,	Adjusted R-squared:  0.0984
F-statistic: 16.26 on 5 and 694 DF,  p-value: 3.496e-15
```

Residuals



Bayesian

```
> summary(regOutp1mcmc1)
Bayes factor analysis
--------------
[1] PctChildPoverty + PctFreeMeal + PctFamilyPoverty + Enrolled + TotalSchools : 784436210286 ±0%

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

```
 Model:
Type: BFlinearModel, JZS
D_Comp ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty + Enrolled + TotalSchools
Data types:
PctChildPoverty :  continuous
PctFreeMeal :  continuous
PctFamilyPoverty :  continuous
Enrolled :  continuous
TotalSchools :  continuous

> summary(regOutp1mcmc)

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                0.9413609 8.519e-03 8.519e-05      8.519e-05
PctChildPoverty   0.0014741 1.415e-03 1.415e-05      1.434e-05
PctFreeMeal      -0.0006744 5.214e-04 5.214e-06      5.214e-06
PctFamilyPoverty -0.0027544 2.040e-03 2.040e-05      2.040e-05
Enrolled          0.0002163 3.656e-05 3.656e-07      3.715e-07
TotalSchools     -0.0218904 3.374e-03 3.374e-05      3.447e-05
sig2              0.0497599 2.687e-03 2.687e-05      2.687e-05
g                 0.0599380 5.834e-02 5.834e-04      5.834e-04

2. Quantiles for each variable:

               2.5%        25%        50%        75%      97.5%
mu         0.9245571  0.9356098  0.9413960  0.9470668  0.9578281
%ChildPov -0.0012569  0.0005113  0.0014653  0.0024352  0.0042532
%FreeMeal -0.0016838 -0.0010261 -0.0006732 -0.0003256  0.0003527
%Fam_ Pov -0.0067145 -0.0041441 -0.0027200 -0.0013837  0.0011986
Enrolled   0.0001431  0.0001925  0.0002164  0.0002407  0.0002884
Tot_Sch   -0.0286017 -0.0241423 -0.0219017 -0.0196961 -0.0150621
sig2       0.0448739  0.0478652  0.0496843  0.0515044  0.0553061
g          0.0160905  0.0305858  0.0450248  0.0690008  0.1954456
```

Logistic Regression

```
> summary(chOutp1)

Call:
glm(formula = DistrictComplete ~ PctChildPoverty + PctFreeMeal +
    PctFamilyPoverty + Enrolled + TotalSchools, family = binomial(),
    data = casch)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7150   0.2449   0.2906   0.3453   2.2196

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       3.7482926  0.4658069   8.047 8.49e-16 ***
PctChildPoverty   0.0316670  0.0307557   1.030  0.30318
PctFreeMeal      -0.0151321  0.0109138  -1.387  0.16559
PctFamilyPoverty -0.0463240  0.0393528  -1.177  0.23914
Enrolled          0.0020846  0.0006752   3.088  0.00202 **
TotalSchools     -0.2057911  0.0586442  -3.509  0.00045 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 312.23  on 699  degrees of freedom
Residual deviance: 275.24  on 694  degrees of freedom
AIC: 287.24

Number of Fisher Scoring iterations: 6
```

Problem 6

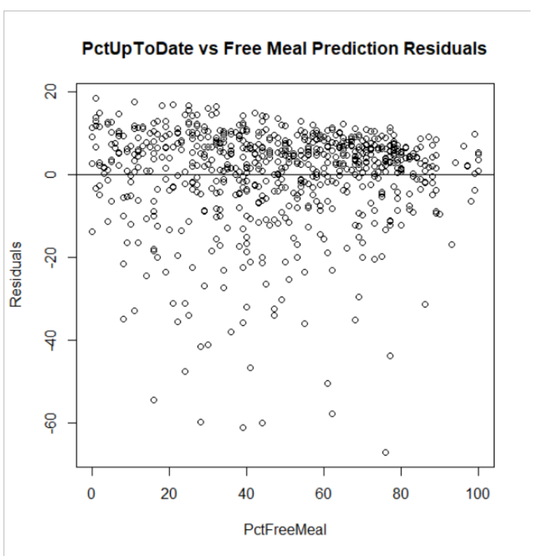
The predictors of up to date vaccines are whether a child receives a free meal as part of the school breakfast or lunch program, if the student is enrolled (or course), is part of the total school population, and has incidence of family poverty. While all measuring significance the free meal and enrollment are most significant as shown below: Bayesian confirmation follows below:

```
> regOutP2

Call:
lm(formula = PctUpToDate ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty +
    Enrolled + TotalSchools, data = casch1)

Coefficients:
```

```
  (Intercept)  PctChildPoverty     PctFreeMeal  PctFamilyPoverty       Enrolled     TotalScho
ols
    82.045848       -0.078131        0.086926         0.284422        0.005684       -0.504669
```

```
> summary(regOutP2)

Call:
lm(formula = PctUpToDate ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty +
    Enrolled + TotalSchools, data = casch1)

Residuals:
    Min      1Q  Median      3Q     Max
-67.110  -3.520   3.195   7.305  18.449

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      82.045848   1.086699  75.500  < 2e-16 ***
PctChildPoverty  -0.078131   0.078780  -0.992  0.32166
PctFreeMeal       0.086926   0.028756   3.023  0.00260 **
PctFamilyPoverty  0.284422   0.112909   2.519  0.01199 *
Enrolled          0.005684   0.002016   2.819  0.00495 **
TotalSchools     -0.504669   0.185808  -2.716  0.00677 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.16 on 694 degrees of freedom
Multiple R-squared:  0.08371,	Adjusted R-squared:  0.0771
F-statistic: 12.68 on 5 and 694 DF,  p-value: 8.313e-12
```

Residuals”



Bayesian

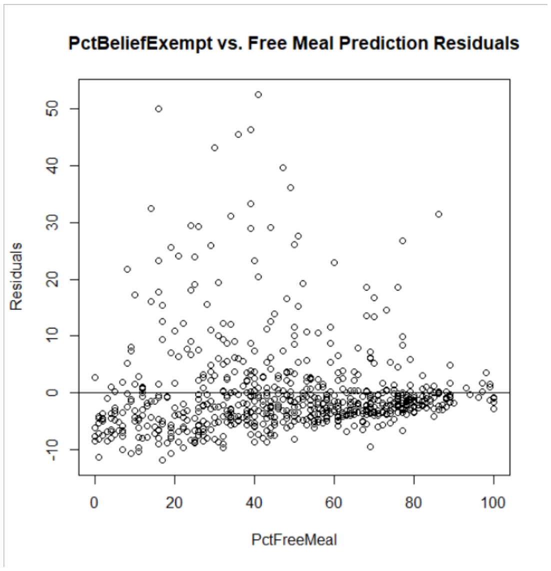
```
> summary(regOutP2mcmc2)
Bayes factor analysis
--------------
[1] PctChildPoverty + PctFreeMeal + PctFamilyPoverty + Enrolled + TotalSchools : 310128100 ±0.01%

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


> summary(regOutP2mcmc)

Iterations = 1:10000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:



                   Mean      SD     Naive SE Time-series SE
Mu                87.53935 0.458956 4.590e-03      4.590e-03
PctChildPoverty   -0.07543 0.077073 7.707e-04      7.707e-04
PctFreeMeal        0.08368 0.028604 2.860e-04      2.860e-04
PctFamilyPoverty   0.27471 0.110743 1.107e-03      1.107e-03
Enrolled           0.00544 0.001988 1.988e-05      2.051e-05
TotalSchools      -0.48317 0.183274 1.833e-03      1.881e-03
sig2             147.83110 7.972515 7.973e-02      7.731e-02
g                  0.05407 0.056540 5.654e-04      5.746e-04

2. Quantiles for each variable:

                2.5%       25%       50%       75%      97.5%
mu          86.649364  87.229861  87.539068  87.848565  88.449612
%ChildPov  -0.225560  -0.127169  -0.075032  -0.023922   0.075636
%FreeMeal   0.028547   0.064463   0.083545   0.102988   0.140023
%FamilyPov  0.056883   0.201392   0.274095   0.347977   0.491752
Enrolled    0.001567   0.004103   0.005414   0.006771   0.009353
Tot_Sch    -0.845933  -0.606350  -0.482311  -0.360686  -0.128349
sig2      133.276262 142.337023 147.462221 152.917268 164.284508
g           0.014556   0.027273   0.040164   0.062237   0.175701
```

Problem 7

The predictors of students with belief exceptions are whether a child received a free meal, if the child and family have incidence of poverty. The results are highlighted below.

```
> regOutP3

Call:
lm(formula = PctBeliefExempt ~ PctChildPoverty + PctFreeMeal +
    PctFamilyPoverty + Enrolled + TotalSchools, data = casch1)

Coefficients:
     (Intercept)  PctChildPoverty      PctFreeMeal  PctFamilyPoverty         Enrolled     TotalScho
ols
       10.039191         0.148355        -0.110112         -0.195207        -0.002484      0.205851

> summary(regOutP3)

Call:
lm(formula = PctBeliefExempt ~ PctChildPoverty + PctFreeMeal +
    PctFamilyPoverty + Enrolled + TotalSchools, data = casch1)

Residuals:
    Min      1Q  Median      3Q     Max
-11.894  -4.102  -2.080   0.748  52.615

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      10.039191   0.739454  13.576  < 2e-16 ***
PctChildPoverty   0.148355   0.053607   2.767   0.0058 **
PctFreeMeal      -0.110112   0.019567  -5.627 2.65e-08 ***
PctFamilyPoverty -0.195207   0.076830  -2.541   0.0113 *
Enrolled         -0.002484   0.001372  -1.811   0.0706 .
TotalSchools      0.205851   0.126434   1.628   0.1040
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.277 on 694 degrees of freedom
Multiple R-squared:  0.1059,	Adjusted R-squared:  0.09945
F-statistic: 16.44 on 5 and 694 DF,  p-value: 2.373e-15
```

Residuals



Bayesian

```
> summary(regOutP3mcmc3)
Bayes factor analysis
--------------
[1] PctChildPoverty + PctFreeMeal + PctFamilyPoverty + Enrolled + TotalSchools : 1.15747e+12 ±0%

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


> summary(regOutP3mcmc)

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                5.775427 0.31357 0.0031357      0.0031357
PctChildPoverty   0.142907 0.05332 0.0005332      0.0005332
PctFreeMeal      -0.106265 0.01943 0.0001943      0.0001943
PctFamilyPoverty -0.188143 0.07607 0.0007607      0.0007607
Enrolled         -0.002436 0.00133 0.0000133      0.0000133
TotalSchools      0.202219 0.12260 0.0012260      0.0012260
sig2             68.538051 3.64982 0.0364982      0.0372755
```

g	0.060543	0.06041	0.0006041	0.0006041

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
mu	5.158400	5.565218	5.77859	5.987978	6.3834281
PctChildPoverty	0.038472	0.106961	0.14219	0.178409	0.2478985
PctFreeMeal	-0.144584	-0.119178	-0.10602	-0.093150	-0.0685137
%FamilyPoverty	-0.334401	-0.239601	-0.18852	-0.137702	-0.0382103
Enrolled	-0.005036	-0.003346	-0.00243	-0.001532	0.0001518
TotalSchools	-0.036209	0.118447	0.20108	0.286480	0.4401417
sig2	61.746629	66.061551	68.41285	70.909892	76.0788191
g	0.016327	0.030471	0.04529	0.069592	0.2007528

Problem 8

Thank you for the opportunity to review predictors of district vaccine completeness, up to date vaccines, and belief exemptions upon vaccine compliance. With an eye upon financial assistance allocation we focused our analysis upon both predictors and validation of our analysis. The results suggest that best predictors of district completeness are whether a child is successfully in TotalSchool and Enrolled. Getting children correctly enrolled into our educational system drives significant results for vaccines and compliance appears to be the result. Using additional measures (called Bayes Factor analysis) show the odds are in favor of this approach. We have shown this graphically.

The analysis of up to date and belief exemptions reveals that the strongest predictor of both is the percentage of students who receive a free meal. This results in terms of significance and additional measure is overwhelming.

Based on the above the results suggest application of financial assistance most strongly towards the free meal program for students and family poverty programs supported by aided by additional efforts to make sure students are properly enrolled and connected to their local school district.

CODE:

```
# IST772 Final Week11  Primrose
dev.off() # Clear the graph window
cat('\014')  # Clear the console
rm(list=ls()) # Clear user objects from the environment
############################################################
# Set the working directory to where you have stored the file
setwd("C:\\Users\\david\\Documents\\Syracuse\\Classes\\IST772Quant\\Week11\\")
load(file = "usVaccines.RData")
# Discovery of dataset
usVaccines
tsVac <- usVaccines
tsVac <- ts(tsVac, frequency = 1, start = c(1980,1))
tsVac
str(tsVac)
summary(tsVac)
plot(tsVac)
cor(tsVac)
plot.ts(tsVac)
plot.ts(tsVac[ ,'DTP1'], main = "DTP1 vs Year")
plot.ts(tsVac[ ,'HepB_BD'], main = "HepB vs Year")
plot.ts(tsVac[ ,'Pol3'], main = "Pol3 vs Year")
plot.ts(tsVac[ ,'Hib3'], main = "Hib3 vs Year")
plot.ts(tsVac[ ,'MCV1'], main = "MCV1 vs Year")
boxplot(tsVac)
summary(tsVac)
tail(tsVac)
plot(diff(tsVac))
dif_usVac <- diff(tsVac)
max(tsVac[ ,'DTP1']) - min(tsVac[ ,'DTP1'])
max(tsVac[ ,"HepB_BD"]) - min(tsVac[ ,"HepB_BD"])
```

```
max(tsVac[ ,'Pol3']) - min(tsVac[ ,'Pol3'])
max(tsVac[ ,'Hib3']) - min(tsVac[ ,'Hib3'])
max(tsVac[ ,'MCV1']) - min(tsVac[ ,'MCV1DTP1'])
sum(tsVac[ ,'DTP1'])
sum(tsVac[ ,"HepB_BD"])
sum(tsVac[ ,'Pol3'])
sum(tsVac[ ,'Hib3'])
sum(tsVac[ ,'MCV1'])
totalVac <- (tsVac[ ,'DTP1'] + tsVac[ ,'HepB_BD'] + tsVac[ ,'Pol3'] + tsVac[ ,'Hib3'] + tsVac[
,'MCV1'])
sum(totalVac)
plot.ts(tsVac[ ,'HepB_DB'], main = "HepB vs Year")
plot.ts(tsVac[ ,'Pol3'], main = "Pol3 vs Year")
plot.ts(tsVac[ ,'Hib3'], main = "Hib3 vs Year")
plot.ts(tsVac[ ,'MCV1'], main = "MCV1 vs Year")
tDTP <- sum(tsVac[ ,'DTP1'])/sum(totalVac)
tHepB <- sum(tsVac[ ,"HepB_BD"])/sum(totalVac)
tPol3 <- sum(tsVac[ ,'Pol3'])/sum(totalVac)
Hib3 <- sum(tsVac[ ,'Hib3'])/sum(totalVac)
MCV1 <- sum(tsVac[ ,'MCV1'])/sum(totalVac)
tDTP
tHepB
tPol3
Hib3
MCV1
View(tsVac)
##################################
# Look at volatility
plot(dif_usVac)
library(changepoint)
cptOut <- cpt.var(dif_usVac, method = "PELT")
```

```
print(cptOut)
plot(cptOut, main="Variance Changepoint Plot")
DPT1 <- usVac[,"DTP1"]
DPT1cp <- cpt.mean(DPT1)
DPT1cp
plot(DPT1cp,cpt.col="grey",cpt.width=5)
HepB_BD <- usVac[,"HepB_BD"]
HepB_BDcp <- cpt.mean(HepB_BD)
HepB_BDcp
plot(HepB_BDcp,cpt.col="grey",cpt.width=5)
Pol3 <- usVac[,"Pol3"]
Pol3cp <- cpt.mean(Pol3)
Pol3cp
plot(Pol3cp,cpt.col="grey",cpt.width=5)
Hib3 <- usVac[,"Hib3"]
Hib3cp <- cpt.mean(Hib3)
Hib3cp
plot(Hib3cp,cpt.col="grey",cpt.width=5)
MCV1 <- usVac[,"MCV1"]
MCV1cp <- cpt.mean(MCV1)
MCV1cp
plot(MCV1cp,cpt.col="grey",cpt.width=5)
# Variance
cpt.var(diff(tsVac)) # Examine the change in variance
cpt.var(diff(tsVac[,"DTP1"]))
cpt.var(diff(usVac[,"HepB_BD"]))
cpt.var(diff(usVac[,"Pol3"]))
cpt.var(diff(usVac[,"Hib3"]))
cpt.var(diff(usVac[,"MCV1"]))
##################################
plot(usVac)
```

```
plot(diff(usVac))
dif_usVac <- diff(usVac)
adf.test(diff(usVac[,"DTP1"]))
adf.test(diff(usVac[,"HepB_BD"]))
adf.test(diff(usVac[,"Pol3"]))
adf.test(diff(usVac[,"Hib3"]))
adf.test(diff(usVac[,"MCV1"]))
acf(diff(usVac[,"DTP1"]))
acf(diff(usVac[,"HepB_BD"]))
acf(diff(usVac[,"Pol3"]))
acf(diff(usVac[,"Hib3"]))
acf(diff(usVac[,"MCV1"]))
cor(diff(usVac))
# Examine stationarity
acf(usVac)
acf(dif_usVac)
############################################
# A different look at volatility (not used but wanted to try it)
plot(dif_usVac)
library(changepoint)
cptOut <- cpt.var(dif_usVac, method = "PELT")
print(cptOut)
plot(cptOut, main="Variance Changepoint Plot")
DPT1 <- usVac[,"DTP1"]
DPT1cp <- cpt.mean(DPT1)
DPT1cp
plot(DPT1cp,cpt.col="grey",cpt.width=5)
HepB_BD <- usVac[,"HepB_BD"]
HepB_BDcp <- cpt.mean(HepB_BD)
HepB_BDcp
plot(HepB_BDcp,cpt.col="grey",cpt.width=5)
```

```
Pol3 <- usVac[,"Pol3"]
Pol3cp <- cpt.mean(Pol3)
Pol3cp
plot(Pol3cp,cpt.col="grey",cpt.width=5)
Hib3 <- usVac[,"Hib3"]
Hib3cp <- cpt.mean(Hib3)
Hib3cp
plot(Hib3cp,cpt.col="grey",cpt.width=5)
MCV1 <- usVac[,"MCV1"]
MCV1cp <- cpt.mean(MCV1)
MCV1cp
plot(MCV1cp,cpt.col="grey",cpt.width=5)

plot(DAXcp,cpt.col="grey",cpt.width=5)
#dev.off()

cpt.var(diff(tsVac)) # Examine the change in variance
cpt.var(diff(tsVac[,"DTP1"]))
cpt.var(diff(usVac[,"HepB_BD"]))
cpt.var(diff(usVac[,"Pol3"]))
cpt.var(diff(usVac[,"Hib3"]))
cpt.var(diff(usVac[,"MCV1"]))

install.packages("bcp")
library(bcp)
bcpDPT1 <- bcp(as.vector(DPT1))
plot(bcpDPT1)
bcpHepB_BD <- bcp(as.vector(HepB_BD))
plot(bcpHepB_BD)
bcpPol3 <- bcp(as.vector(Pol3))
plot(bcpPol3)
```

```
bcpHib3 <- bcp(as.vector(Hib3))
plot(bcpHib3)
bcpDTP1 <- bcp(as.vector(DTP1))
plot(bcpDTP1)
bcpMCV1 <- bcp(as.vector(MCV1))
plot(bcpMCV1)

bcpDAX <- bcp(as.vector(DAX))
plot(bcpDAX)
#png("Figure11_13.png", width = 8, height = 6, units = 'in', res = 300)
plot(bcpDAX,outer.margins = list(left = unit(4,"lines"), bottom = unit(3, "lines"), right = unit(3,
"lines"), top = unit(2,"lines")), main=NULL)
#dev.off()

#png("Figure11_14.png", width = 6, height = 6, units = 'in', res = 300)
plot(bcpDAX$posterior.prob>.95)
#dev.off()

# Examine stationarity
acf(testPanel, main="ACF: Undifferenced Series")
diffTestPanel <- diff(testPanel)
acf(diffTestPanel, main="ACF: Differenced Series")
require(tseries)
adf.test(diffTestPanel)

# Look at volatility
plot(diffTestPanel, main="Differenced Series")
print("Overall SD of differenced series:")
print(sd(diffTestPanel))
library(changepoint)
cptOut <- cpt.var(diffTestPanel, method = "PELT")
```

```
print(cptOut)
plot(cptOut, main="Variance Changepoint Plot")
################################################



#########################################################
#  Public Schools
#########################################################
# Set the working directory to where you have stored the file
setwd("C:\\Users\\david\\Documents\\Syracuse\\Classes\\IST772Quant\\Week11\\")
load(file = "allSchoolsReportStatus.RData")
allSchoolsReportStatus
ASRS <- allSchoolsReportStatus
dim(ASRS)
class(ASRS)
View(ASRS)
str(ASRS)
summary(ASRS)
table(ASRS)
colnames(ASRS)
head(ASRS)
table(ASRS$pubpriv, ASRS$reported)
perc <- table(ASRS$pubpriv, ASRS$reported)/7381
dmargin <-margin.table(as.matrix(table(ASRS$pubpriv, ASRS$reported)))
dmargin
table(ASRS$reported)
table(ASRS$pubpriv)

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

```
toast <- matrix(c(252,1397,148,5584),ncol=2,byrow=TRUE) # Create a two column structure
using the matrix() command
colnames(toast) <- c("NO","YES") # Label the columns
rownames(toast) <- c("PRIVATE","PUBLIC") # Label the rows
toast <- as.table(toast) # Convert from metric to table
toast # Show the table on the console
margin.table(toast) # This is the grand total of toast drops
margin.table(toast,1) # These are the marginal totals for rows
margin.table(toast,2) # These are the marginal totals for columns
toastProbs <- toast/margin.table(toast) # Calculate probabilities
toastProbs # Report probabilities to console
barplot(toastProbs)
barplot(toast)
hist(toast)
barplot(cumsum(toastProbs))
cumsum(toast)
margin.table(toast,2)
margin.table(toast)
USMar <- margin.table(toast,2)/margin.table(toast)
USMar
#####################################################

#####################################################
#Question 3
#####################################################
# Set the working directory to where you have stored the file
setwd("C:\\Users\\david\\Documents\\Syracuse\\Classes\\IST772Quant\\Week11\\")
#"C:\Users\david\Documents\Syracuse\Classes\IST772Quant\Week11\districts17.RData"
load(file = "districts17.RData")
districts
```

```
install.packages("xlsx")
library(xlsx)
write.xlsx(districts,
"C:\\Users\\david\\Documents\\Syracuse\\Classes\\IST772Quant\\Week11\\mydata_dist.xlsx")
View(districts)
sum(districts$TotalSchools)
casch <- districts
dim(casch)
class(casch)
View(casch)
str(casch)
summary(casch)
table(casch)
colnames(casch)
head(casch)
sum(casch$PctUpToDate)
sum(casch$Enrolled)
sum(casch$WithoutDTP)
sum(casch$WithoutPolio)
sum(casch$WithoutMMR)
sum(casch$WithoutHepB)
sum(casch$WithoutDTP)
DTP <- 1-(sum(casch$WithoutDTP)/sum(casch$Enrolled))
Polio <- 1-(sum(casch$WithoutPolio)/sum(casch$Enrolled))
MMR <- 1-(sum(casch$WithoutMMR)/sum(casch$Enrolled))
HepB <-1-(sum(casch$WithoutHepB)/sum(casch$Enrolled))
DTP
Polio
MMR
HepB
casch$D_Comp <- as.numeric(casch$DistrictComplete)
```

```
casch
names(casch)
casch$D_Comp
names(casch)
casch$D_Comp <- NULL
names(casch)
casch1 <- casch
casch1$DistrictName <- NULL
casch1$DistrictComplete <- NULL
casch1$WithoutDTP <- NULL
casch1$WithoutPolio <- NULL
casch1$WithoutMMR  <- NULL
casch1$WithoutHepB <- NULL
names(casch1)
cor(casch1)
class(casch1$TotalSchools)
names(casch1)
casch1

##############################################
#Linear Regressions and Residual plots
regOutp1 <- lm(D_Comp ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty + Enrolled +
TotalSchools, data=casch1)
summary(regOutp1)
regOutp1.res = resid(regOutp1)
plot(casch1$TotalSchools, regOutp1.res, ylab="Residuals", xlab="TotalSchools", main="District
Complete vs. TS Prediction Residuals")
abline(0, 0)
regOutP2 <- lm(PctUpToDate ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty + Enrolled
+ TotalSchools, data=casch1)
#regOutP2
```

```
summary(regOutP2)
regOutP2.res = resid(regOutP2)
plot(casch1$PctFreeMeal, regOutP2.res, ylab="Residuals", xlab="PctFreeMeal",
main="PctUpToDate vs Free Meal Prediction Residuals")
abline(0, 0)
regOutP3 <- lm(PctBeliefExempt ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty +
Enrolled + TotalSchools, data=casch1)
#regOutP3
summary(regOutP3)
regOutP3.res = resid(regOutP3)
plot(casch1$PctFreeMeal, regOutP3.res, ylab="Residuals", xlab="PctFreeMeal",
main="PctBeliefExempt vs. Free Meal Prediction Residuals")
abline(0, 0)
##############################################
##############################################
#Bayesian Regressions
regOutp1mcmc <- lmBF(D_Comp ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty +
Enrolled + TotalSchools, data=casch1, posterior=TRUE, iterations=10000)
regOutp1mcmc
summary(regOutp1mcmc)
regOutp1mcmc1 <- lmBF(D_Comp ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty +
Enrolled + TotalSchools, data=casch1)
#regOutp1mcmc1
summary(regOutp1mcmc1)
regOutP2mcmc <- lmBF(PctUpToDate ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty +
Enrolled + TotalSchools, data=casch1, posterior=TRUE, iterations=10000)
#regOutP2mcmc
summary(regOutP2mcmc)
regOutP2mcmc2 <- lmBF(PctUpToDate ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty
+ Enrolled + TotalSchools, data=casch1)
#regOutP2mcmc2
```

```
summary(regOutP2mcmc2)
regOutP3mcmc <- lmBF(PctBeliefExempt ~ PctChildPoverty + PctFreeMeal +
PctFamilyPoverty + Enrolled + TotalSchools, data=casch1, posterior=TRUE, iterations=10000)
#regOutP3mcmc
summary(regOutP3mcmc)
regOutP3mcmc3 <- lmBF(PctBeliefExempt ~ PctChildPoverty + PctFreeMeal +
PctFamilyPoverty + Enrolled + TotalSchools, data=casch1)
#regOutP3mcmc3
summary(regOutP3mcmc3)
##############################################
# Looking for R-square but couldn't get this working...
rsqList <- 1 - (regOutp1mcmc1[ ,"sig2"] / var(casch1$D_Comp))
mean(rsqList)                    # Overall mean R-squared
quantile(rsqList,c(0.025))
quantile(rsqList,c(0.975))

rsqList <- 1 - (regOutP2mcmc2[,"sig2"] / var(casch1$PctUpToDate))
mean(rsqList)                    # Overall mean R-squared
quantile(rsqList,c(0.025))
quantile(rsqList,c(0.975))

rsqList <- 1 - (regOutP3mcmc[,"sig2"] / var(casch1$PCtBeliefExempt))
mean(rsqList)                    # Overall mean R-squared
quantile(rsqList,c(0.025))
quantile(rsqList,c(0.975))
##############################################
##############################################
# Logitic Regression on categorical District Complete
chOutp1 <- glm(formula = DistrictComplete ~ PctChildPoverty + PctFreeMeal +
PctFamilyPoverty + Enrolled + TotalSchools, family = binomial(),data=casch)
#chOutp1
```

```
summary(chOutp1)
casch$DistrictName <- NULL
chOutp2 <- glm(formula = PctUpToDate ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty
+ Enrolled + TotalSchools, family = binomial(),data=casch)
chOutp2 <- glm(formula = PctUpToDate ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty
+ Enrolled + TotalSchools, family = binomial(),data=casch)
#chOutP2
summary(chOutP2)
chOutP3 <- lm(PctBeliefExempt ~ PctChildPoverty + PctFreeMeal + PctFamilyPoverty +
Enrolled + TotalSchools, data=casch)
#chOutP3
```