STAT512 Team4

COVID-19 Infection and Death Cases in Indiana

Yuan Gao, Xiaoke He, Baoxuan Tao, Hao Yin

Introduction

- COVID: Dec. 2019
    - economics
    - infections
    - deaths
- Masks: Jul. 2020-Apr. 2021
- Vaccine: Dec. 2020
- Booster: Aug. 2021

Introduction

2-Dose Vaccine

- Study has shown an efficacy of 95% against infection.

Booster

- Study has shown an efficacy of 96 % against severe illness and death.

Masks

- Study has shown that masks block 80%-90% of virus.

Introduction

- Efficacy: measured in lab

- Effectiveness: social study

- Effectiveness of Vaccine, Booster, Mask

- Confirmatory Study

- Weekly Data

- Indiana

Measurement

Source of Variables

- All data comes from Centers for Disease Control and Prevention. 6 continuous variables and 1 categorical variables are in the dataset.

- Continuous variables are Total Vaccine, First Dose, Complete Series, Booster, Infection, and Death.

- Categorical variable is Mask.

- All 6 continuous variables record data in Indiana.

Data Cleaning

- The original vaccine data is the total number of administered vaccines per day.

- To match the unit of infection and death variables, we calculate the weekly total vaccine number by summing up every day's vaccine in one week.

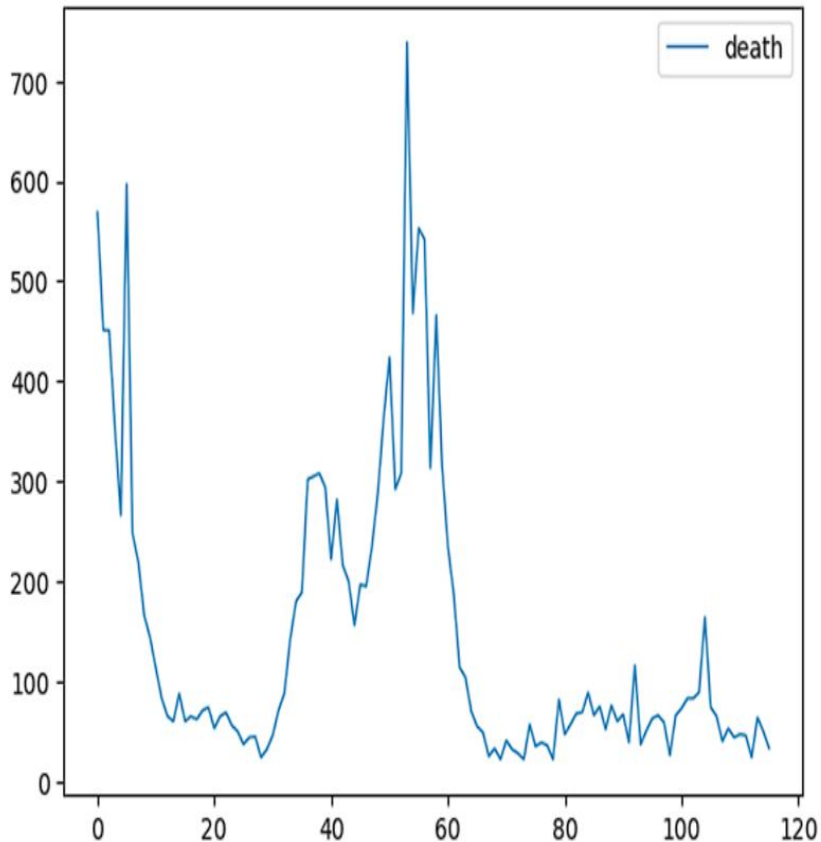- In order to eliminate outliers, we delete all the data containing 0.

Measurement

Variable Description
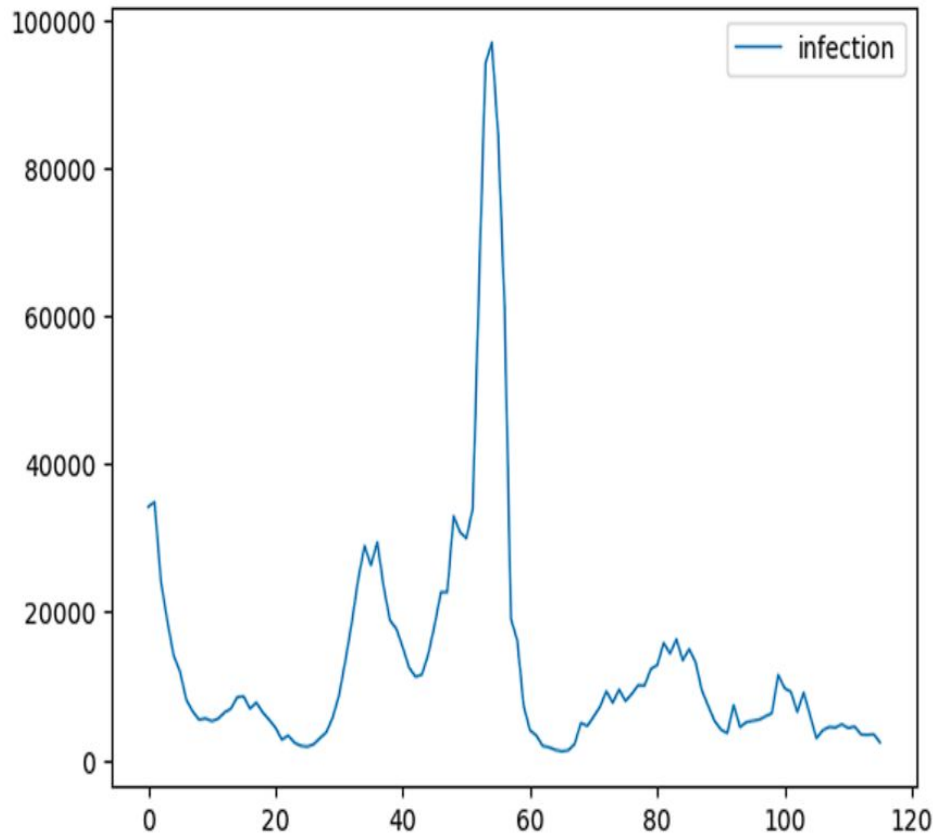
- Infection records the weekly infection cases in Indiana from December 31, 2020 to March 29, 2023.
- As vaccine takes 1-2 week to take effect, we postpone the infection cases two weeks to correspond to the effective vaccine.

PURDUE UNIVERSITY®

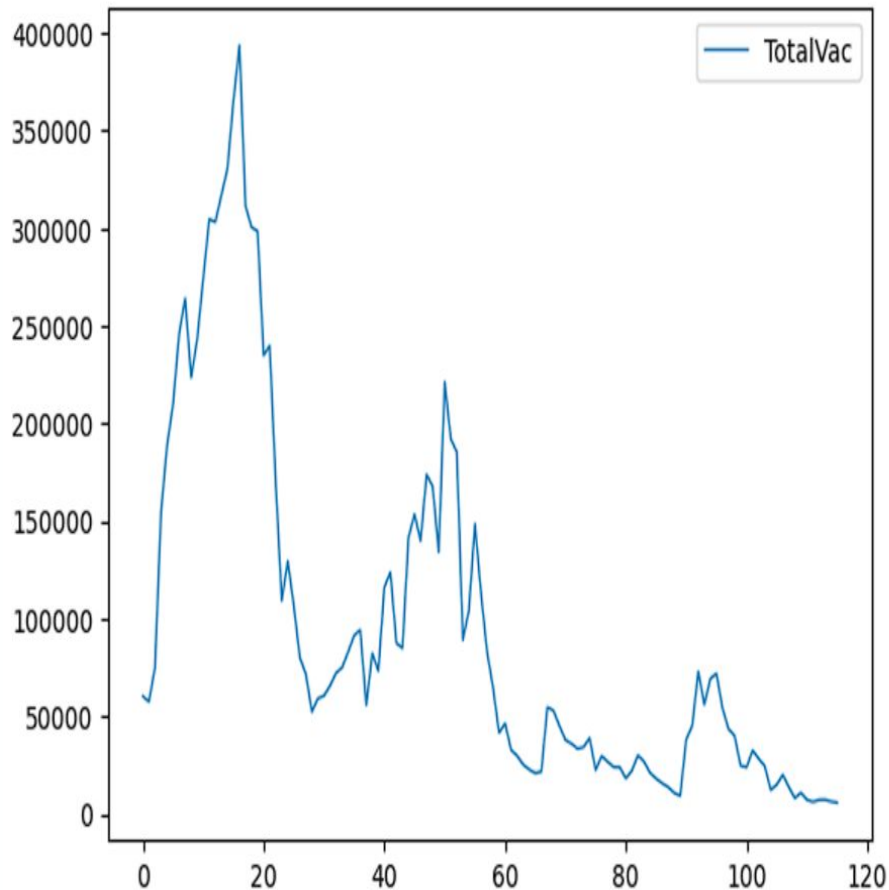# *Measurement*

Variable Description

- Death records the weekly death cases in Indiana from December 31, 2020 to March 29, 2023.

- Similar to Infection variables, we postpone the death cases by two weeks to correspond to the effective vaccine.

PURDUE UNIVERSITY.

Measurement

Variable Description

- Total Vaccine is the total number of administered vaccines per week in Indiana, which includes the first dose, second dose, and both first and second booster.
- Different providers like Pfizer-BioNTech, Moderna, Novavax, and Johnson & Johnson.
- CDC's data may overestimate the total number of vaccines.
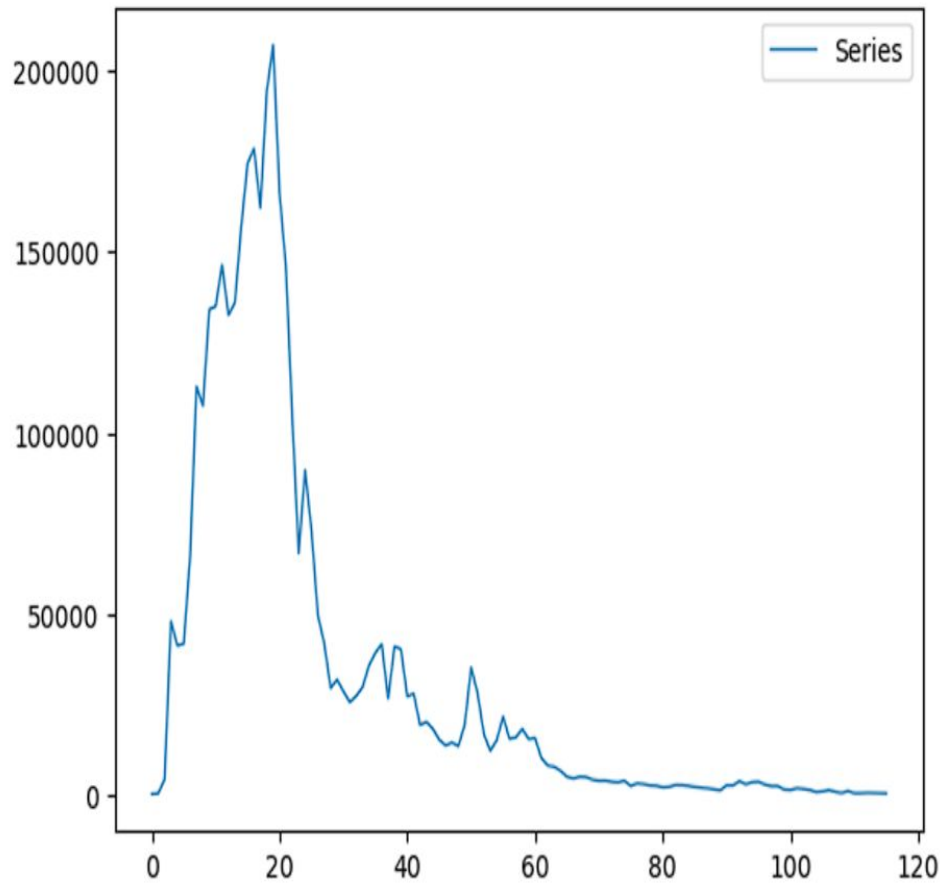
Measurement

Variable Description
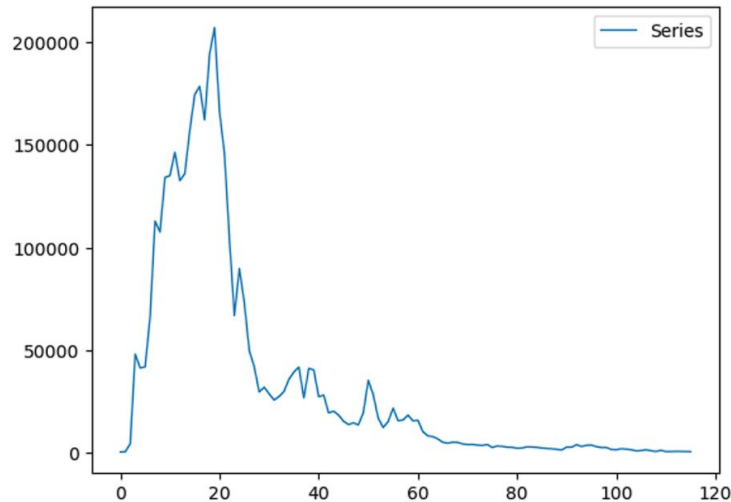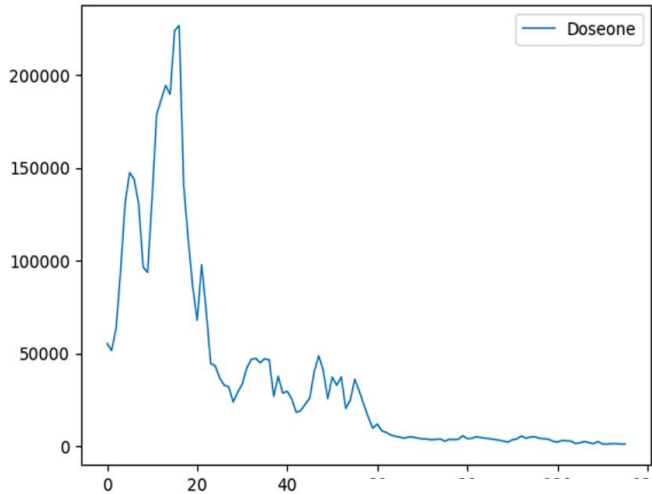
- The Doseone and Series are the total number of the first-dose vaccine and second dose in Indiana per week from different providers.

- CDC's data may overestimate the total number of first dose vaccine but underestimate the total number of second dose vaccine.
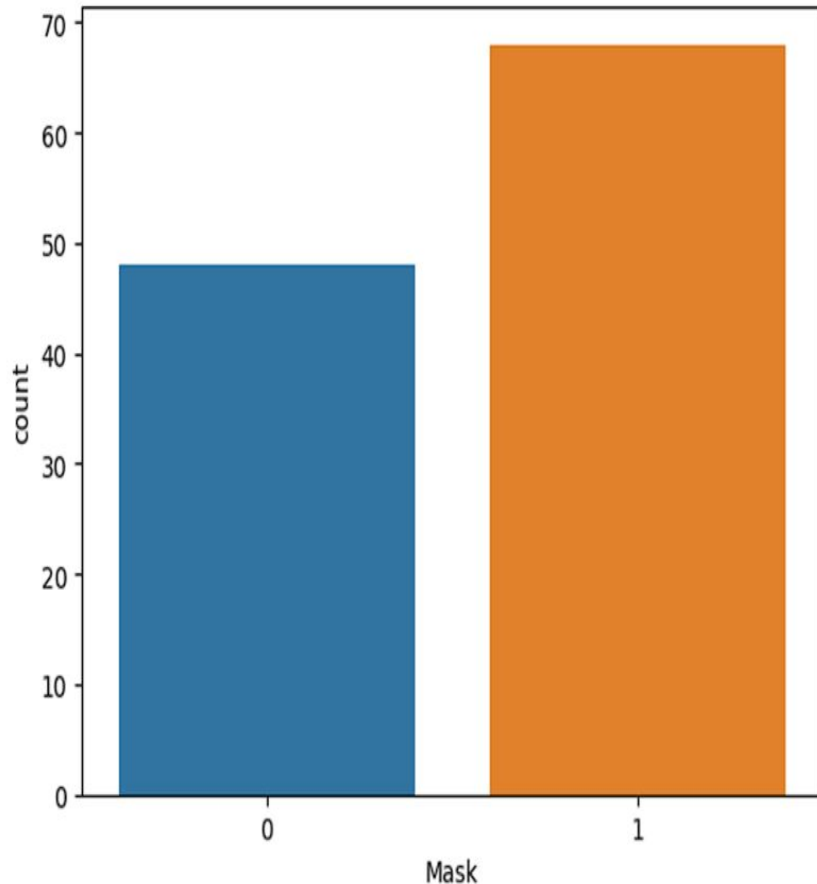
Measurement

Variable Description

- Booster is the total number of first booster dose vaccines in Indiana per week from different providers.

Measurement

Variable Description

- The mask is the categorical variable which includes 0 and 1.
- If Mask is 0, it means the mask mandate policy ended in Indiana. If Mask is 1, it means the mask mandate policy was still activated.
- The mask mandate policy required everyone over the age of 8 to wear a mask both indoors and outside wherever social distancing is not possible since July 27, 2020. The mask mandate ended on April 6, 2021. (IN gov, July 27)

Research Question 1&2

1. Do complete series have significant impact on infected cases, given death cases and first dose ?

  H0: beta3 = 0 given beta2 and beta5 Ha: beta3 is not 0 given beta2 and beta5
  Reduced model: Y = beta0 + beta5(X5) + beta2(X2)
  Full model: Y = beta0 + beta5(X5) + beta2(X2) + beta3(X3)

2. Do complete series and first dose have the same impact on infected cases, given death cases?

  H0: beta2 = beta3 = betanew Ha: beta2 is not equal to beta3
  Reduced model: Y = beta0 + betanew(X2 + X3) + beta5(X5)
  Full model: Y = beta0 + beta2(X2) + beta3(X3) + beta5(X5)

X2: first dose
X3: complete series
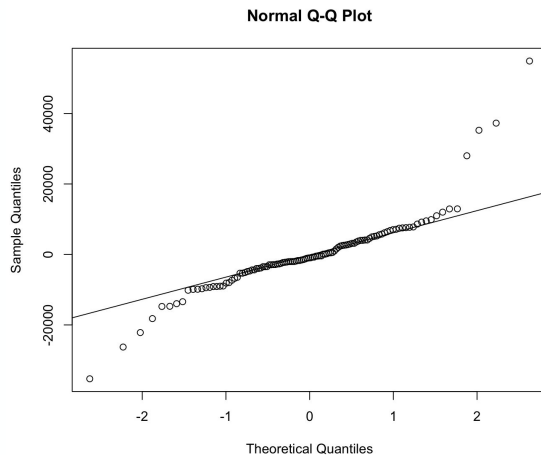X5: death cases
Y: infection cases

PURDUE
UNIVERSITY®

Research Question 1&2 Diagnostics

Shapiro-Wilk normality test

data: res
W = 0.85227, p-value = 2.153e-09

studentized Breusch-Pagan test

data: model
BP = 35.205, df = 3, p-value = 1.103e-07

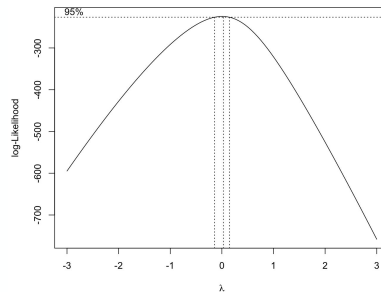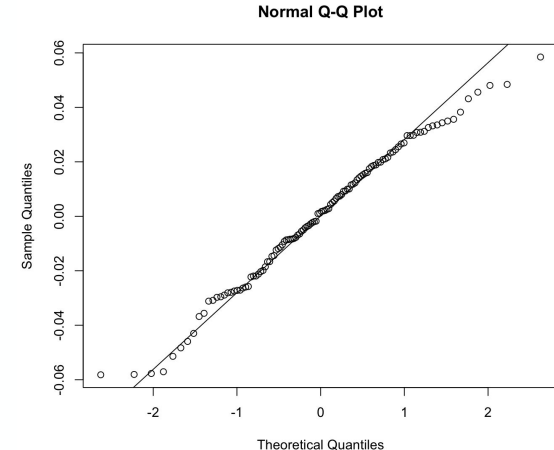**Log transformation on X, Box-cox transformation on Y(λ = 0.0303)**

Shapiro-Wilk normality test

data: residuals(modelNew)
W = 0.98693, p-value = 0.327

studentized Breusch-Pagan test

data: modelNew
BP = 7.5565, df = 3, p-value = 0.05612

Research Question 1&2 Diagnostics

Added-Variable Plots
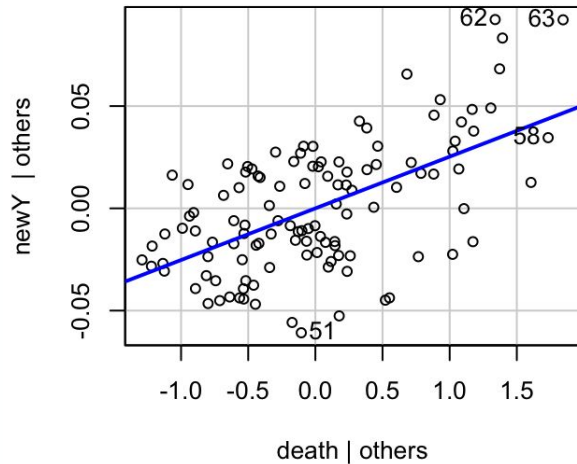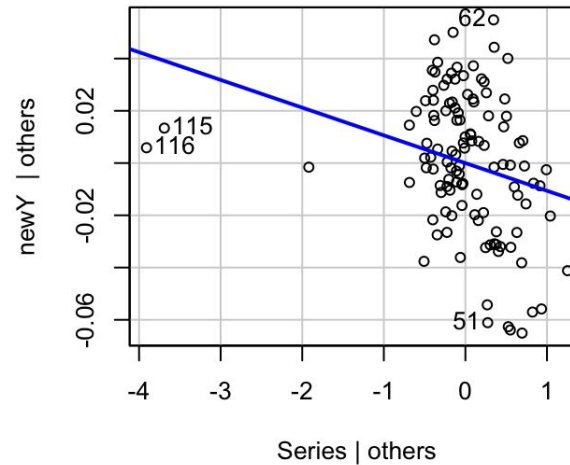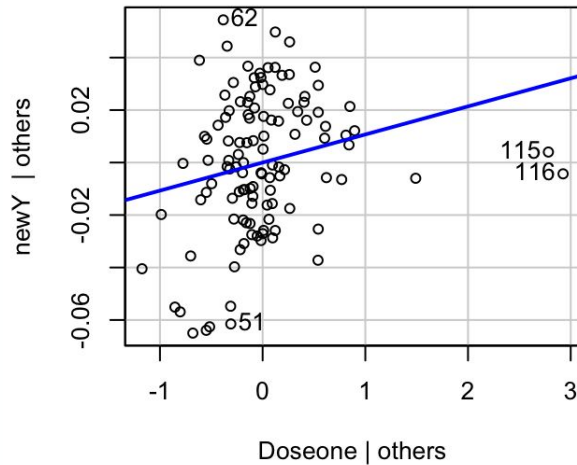
All marginal effects of the predictors are significant.

Research Question 1&2 Diagnostics

```
> sum(abs(rstudent(modelNew)) > 2)
[1] 0
> infl = lm.influence(modelNew)$hat
> length(which(infl[] > 2 * (4 / 116)))
[1] 3
```

```
> sum(d[which(abs(d[, 2]) > 1 & abs(d[, 3])> 1 & abs(d[, 4]) > 1)])
[1] 0
> length(dff[dff > 1])
[1] 0
> sum(Cooksdistance > minor)
[1] 0
> sum(Cooksdistance > major)
[1] 0
```

0 Y outliers, 3 x outliers

No influential points
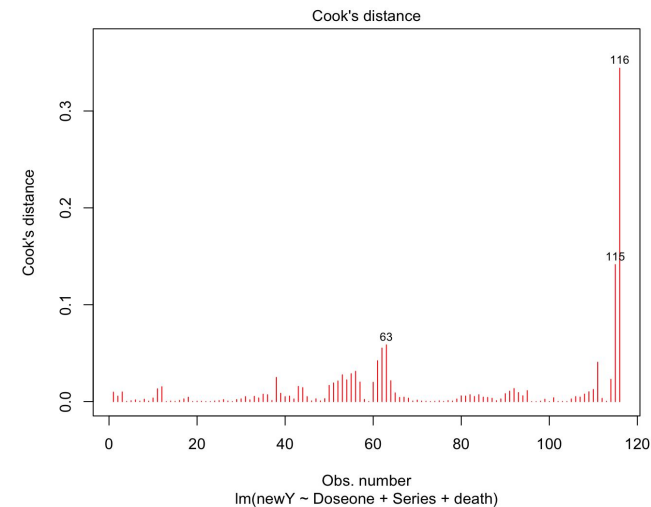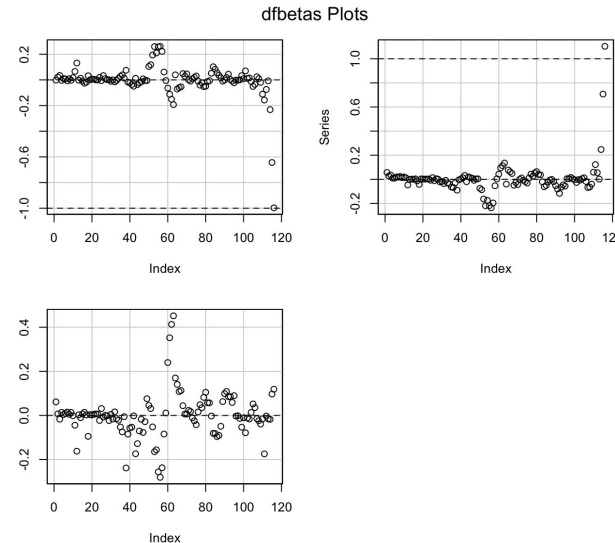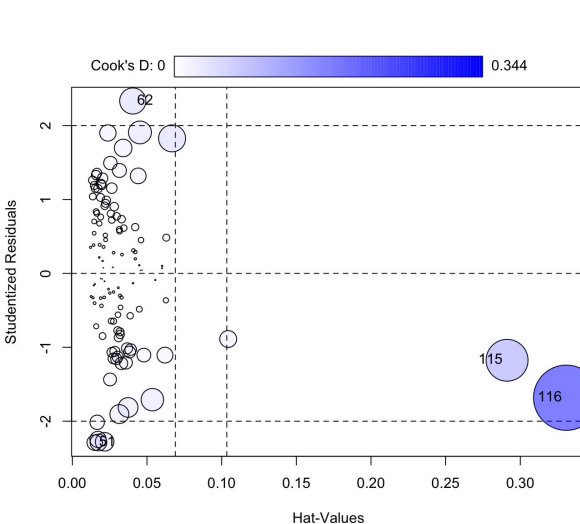
Research Question 1&2 Diagnostics
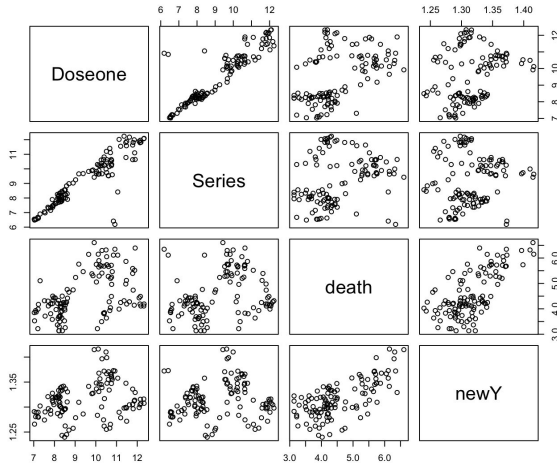
> VIF(lm(Doseone~Series+death, new))
[1] 7.283964
> VIF(lm(Series~Doseone+death, new))
[1] 6.085945
> VIF(lm(death~Doseone+Series, new))
[1] 1.531526

Multicollinearity

Ridge Trace Plot
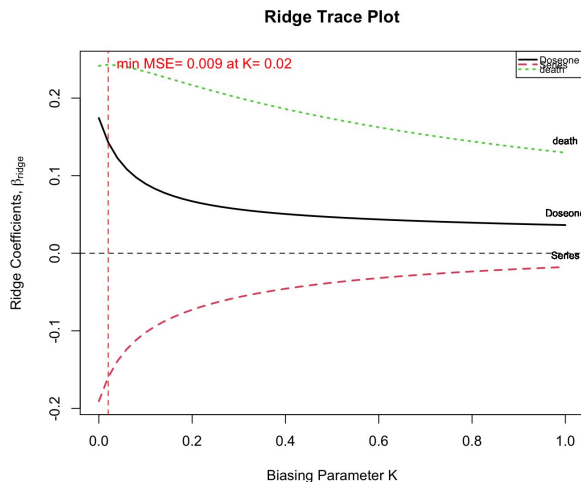
	Doseone	Series	death
k=0	7.28396	6.08595	1.53153
k=0.02	4.65084	3.95688	1.31869
k=0.04	3.25454	2.82358	1.18521
k=0.06	2.42528	2.14714	1.08976
k=0.08	1.89220	1.70962	1.01549
k=0.1	1.52885	1.40920	0.95437
k=0.12	1.26971	1.19316	0.90216
k=0.14	1.07811	1.03194	0.85640
k=0.16	0.93221	0.90792	0.81555
k=0.18	0.81835	0.81007	0.77862

Ridge Regression

Research Question 1&2

Call:
lm(formula = newY ~ Doseone + Series + death, data = new)

Residuals:
```
     Min        1Q    Median        3Q       Max
-0.058194 -0.018880  0.001447  0.019094  0.058479
```

Coefficients:
```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.196178   0.016614  71.998  < 2e-16 ***
Doseone     0.010701   0.004322   2.476  0.01478 *
Series     -0.010609   0.003580  -2.964  0.00371 **
death       0.025270   0.003377   7.483 1.77e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.02608 on 112 degrees of freedom
Multiple R-squared: 0.5122,   Adjusted R-squared: 0.4991
F-statistic: 39.19 on 3 and 112 DF, p-value: < 2.2e-16

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 100 bootstrap replicates

CALL :
boot.ci(boot.out = modelrid, type = "perc", index = 2)

Intervals :
```
Level      Percentile
95%   ( 0.0022,  0.0077 )
```
Calculations and Intervals on Original Scale
Some percentile intervals may be unstable

Call:
lmridge.default(formula = newY ~ Doseone + Series + death, data = new,
    K = 0.14)

Coefficients: for Ridge parameter K= 0.14
```
            Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept   1.2066         0.3447      0.3763       0.9162   0.3615
Doseone     0.0048         0.0779      0.0274       2.8452   0.0053 **
Series     -0.0049        -0.0879      0.0268      -3.2780   0.0014 **
death       0.0238         0.2272      0.0244       9.3040   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ridge Summary
```
      R2     adj-R2   DF ridge        F       AIC        BIC
 0.37720    0.36610    2.13765 38.30602 -841.63815 -284.33547
```
Ridge minimum MSE= 0.02213321 at K= 0.14
P-value for F-test (2.13765 , 113.4471) = 5.317365e-14

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 100 bootstrap replicates

CALL :
boot.ci(boot.out = modelrid, type = "perc", index = 3)

Intervals :
```
Level      Percentile
95%   (-0.0067, -0.0030 )
```
Calculations and Intervals on Original Scale
Some percentile intervals may be unstable

beta2 > 0 ???

1. The amount of data is insufficient and limited to the state of Indiana.
2. It is uncertain whether a two-week delay is enough for the first dose to manifest.
3. Covid-19 virus is notorious for its fast mutation, so it may be the case that the virus has already mutated by the time that dose one is taken.

PURDUE UNIVERSITY

Research Question 1&2

```
> FS = MSR / MSE
> FS
[1] 7.627347
> qf(1 - 0.05, reduced$df.residual -
full$df.residual, full$df.residual)
[1] 3.925834
> p = 1 - pf(FS, reduced$df.residual -
full$df.residual, full$df.residual)
> p
[1] 0.006720598

> FS = MSR/MSE
> FS
[1] 8.78373
> qf(0.95,
reduced$df.residual-full$df.residual,
full$df.residual)
[1] 3.925834
> p =
1-pf(FS,reduced$df.residual-full$df.resi
dual, full$df.residual)
> p
[1] 0.003713453
```

Question 1 conclusion: Complete series has significant impact on infected cases, given death cases and first dose.

+

Question 2 conclusion: Complete series do not have the same impact on infected cases as the first dose, given death cases.

→

A full series of Covid-19 vaccines can restrict the infection rate, which may indirectly prove that a full series vaccination is able to increase the immunity to Covid-19.

Research Question 1&2

K-fold Cross Validation

nvmax	RMSE	Rsquared	MAE	RMSESD	RsquaredSD	MAESD
1	4	0.02722844	0.487252	0.02208553	0.002404326	0.1818767	0.002062165

The model behaves stably across different data and has good predictive power

Research Question 4

Does complete series have the same impact as the booster shot with the first dose and death cases in the model?

X2: first dose
X3: complete series
X4: booster
X5: death cases
Y: infection cases

Ho: beta3 = beta4
Reduced Model: Y = beta2X2+betanew(X3+X4)+beta5X5
Ha: beta3 ≠ beta4
Full Model: Y = beta2X2+beta3X3+beta4X4+beta5X5

```
Call:
lm(formula = y ~ x2 + x3 + x4 + x5)

Residuals:
   Min    1Q Median    3Q    Max
-27566 -3442   -722  4634  43276

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -107.21121 1538.69340  -0.070 0.944632
x2             0.85508    0.21243   4.025 0.000132 ***
x3            -1.09480    0.22240  -4.923 4.75e-06 ***
x4            -0.04840    0.04735  -1.022 0.309829
x5           111.78220   11.09195  10.078 1.03e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9634 on 77 degrees of freedom
Multiple R-squared:  0.7536,    Adjusted R-squared:  0.7408
F-statistic: 58.87 on 4 and 77 DF,  p-value: < 2.2e-16
```

Research Question 4

Does complete series have the same impact as the booster shot with the first dose and death cases in the model?

- BP test
- Shapiro test
- qq plot
- outlier

- non constant error variance
- non normal error
- 6 outliers

```
> # Check Variance Constancy
> bptest(covidFull.mod)

        studentized Breusch-Pagan test

data:  covidFull.mod
BP = 23.856, df = 4, p-value = 8.538e-05

> # Check Variance Normality
> shapiro.test(residuals(covidFull.mod))

        Shapiro-Wilk normality test

data:  residuals(covidFull.mod)
W = 0.89307, p-value = 4.981e-06

> # Check Outlier
> data$outlier <- abs(rstudent(covidFull.mod)) > 2
> sum(data$outlier)
[1] 6
```
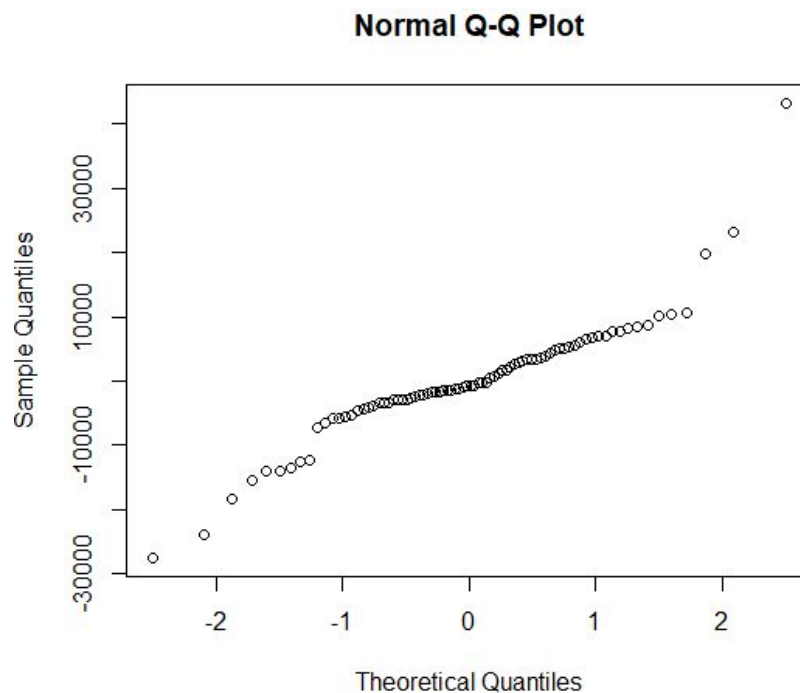
Normal Q-Q Plot

Sample Quantiles (y-axis), Theoretical Quantiles (x-axis)

PURDUE
UNIVERSITY.

Research Question 4

Does complete series have the same impact as the booster shot with the first dose and death cases in the model?

- X Transformation
- Box-Cox Y Transformation
- BP test
- Shapiro test
- qq plot

After transformation:
- constant error variance
- normal error


```
> # Check Constant Variance
> bptest(covidTrans.mod)

        studentized Breusch-Pagan test

data:  covidTrans.mod
BP = 9.2966, df = 4, p-value = 0.0541

> # Check Normality
> shapiro.test(residuals(covidTrans.mod))

        Shapiro-Wilk normality test

data:  residuals(covidTrans.mod)
W = 0.98985, p-value = 0.811

> qqnorm(residuals(covidTrans.mod))
```
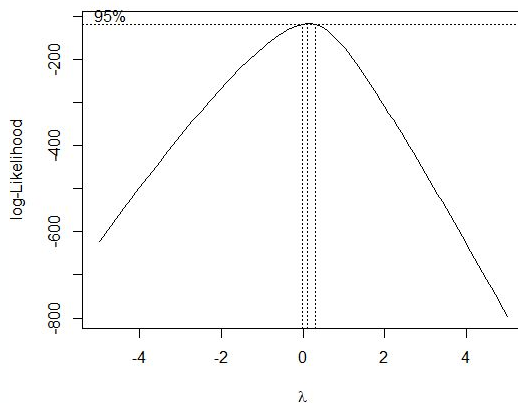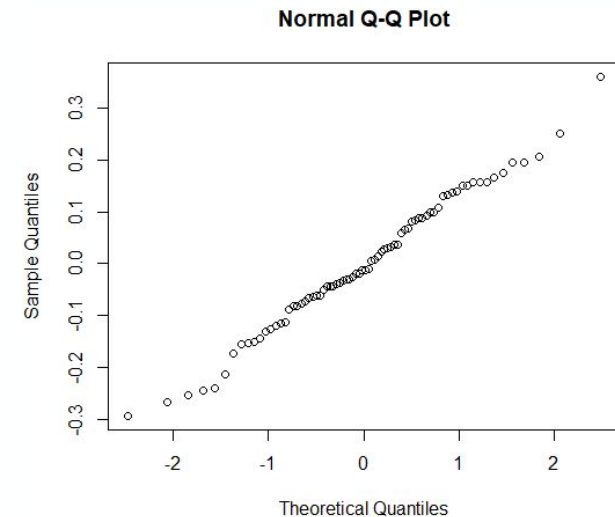

Normal Q-Q Plot

Research Question 4

Does complete series have the same impact as the booster shot with the first dose and death cases in the model?

- Added Variable Plots
- DFBETAS, DFFITS, Cook's Distance

- no 0 marginal effect
- few influential points

Added-Variable Plots

```
> sum(d[which(abs(d[, 2]) > 1 & abs(d[, 3])> 1 & abs(d[, 
[1] 0
> dff = dffits(covidTrans.mod)
> length(dff[dff > 1])
[1] 1
> minor = qf(0.2, df1 = 4, df2 = 76 - 4)
> major = qf(0.5, df1 = 4, df2 = 76 - 4)
> Cooksdistance = cooks.distance(covidTrans.mod)
> sum(Cooksdistance > minor)
[1] 0
> sum(Cooksdistance > major)
[1] 0
```
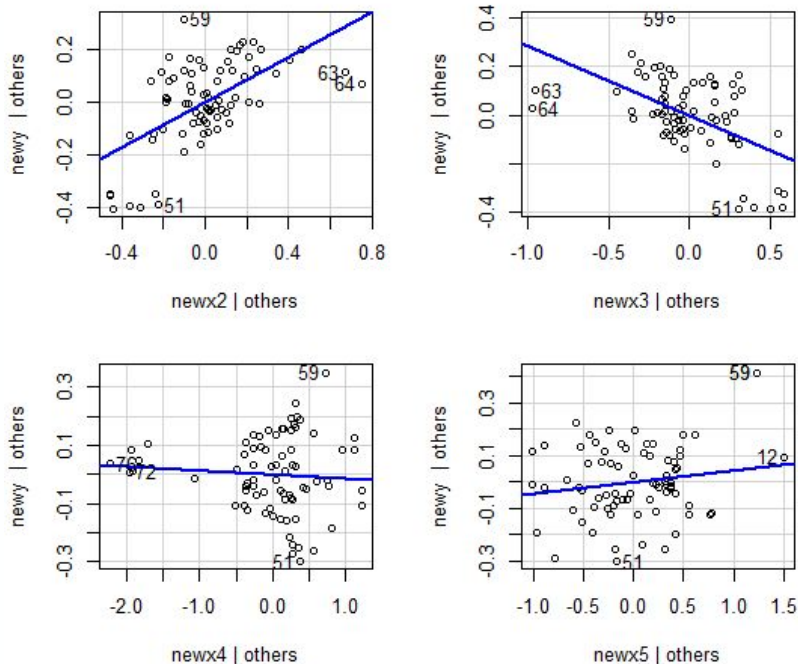
Research Question 4

Does complete series have the same impact as the booster shot with the first dose and death cases in the model?

- pairs plot
- VIF

- no 0 marginal effect
- few influential points
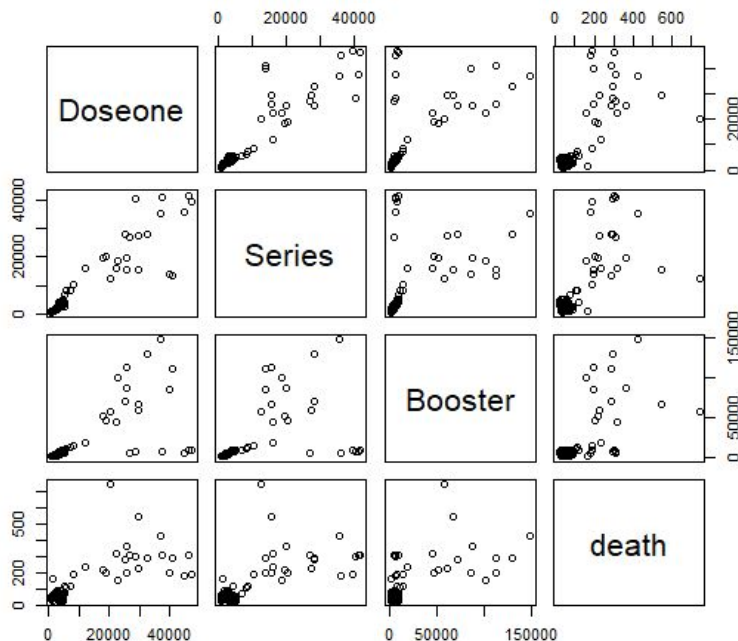

```
> VIF(lm(Doseone~Series+Booster+death, data=data))
[1] 9.202451
> VIF(lm(Series~Doseone+Booster+death, data=data))
[1] 7.146584
> VIF(lm(Booster~Doseone+Series+death, data=data))
[1] 2.143072
> VIF(lm(death~Doseone+Series+Booster, data=data))
[1] 2.385914
```
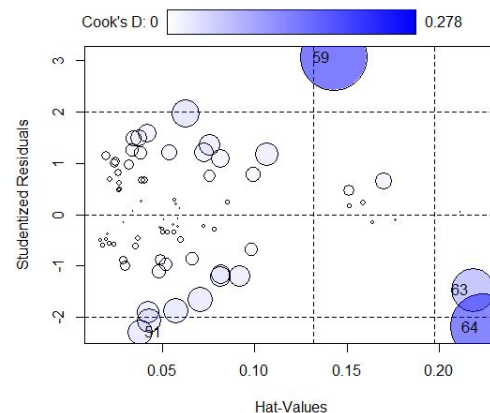
PURDUE UNIVERSITY.

Research Question 4

Does complete series have the same impact as the booster shot with the first dose and death cases in the model?

- Ridge Regression
- multicollinearity removed

```
> vif(Ridge.mod)
          newx2    newx3   newx4   newx5
k=0     23.34991 16.96300 3.08289 2.96793
k=0.02   7.77698  6.19320 2.60475 2.33742
k=0.04   3.99406  3.49207 2.26019 2.02151
k=0.06   2.49969  2.37503 1.98702 1.79395
k=0.08   1.75025  1.78400 1.76388 1.61221
k=0.1    1.31640  1.42205 1.57839 1.46094
k=0.12   1.03997  1.17829 1.42217 1.33226
k=0.14   0.85136  1.00301 1.28921 1.22129
k=0.16   0.71594  0.87088 1.17498 1.12460
k=0.18   0.61479  0.76772 1.07606 1.03969
k=0.2    0.53683  0.68496 0.98978 0.96462
```

Ridge Trace Plot
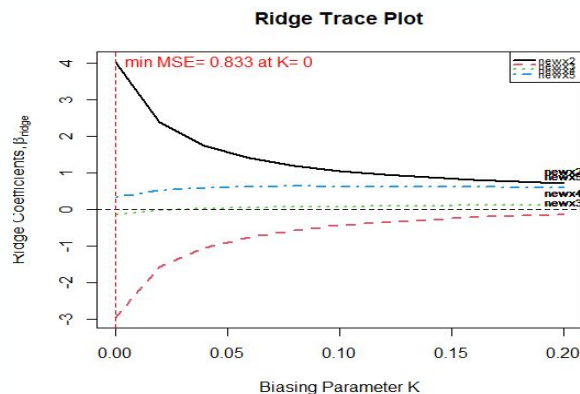

```
> summary(covidRidge.mod)

Call:
lmridge.default(formula = newy ~ newx2 + newx3 + newx4 + newx5,
    data = newdata, K = 0.12)


Coefficients: for Ridge parameter K= 0.12
          Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept   1.4229       -6.4575      2.6821      -2.4076   0.0186
newx2       0.1015        0.9483      0.1570       6.0422   <2e-16
newx3      -0.0337       -0.3469      0.1671      -2.0763   0.0414
newx4       0.0085        0.0924      0.1835       0.5035   0.6161
newx5       0.0845        0.6244      0.1776       3.5148   0.0008

Intercept *
newx2     ***
newx3     *
newx4
newx5     ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge Summary
      R2     adj-R2    DF ridge         F       AIC       BIC
   0.43320   0.40960    2.53802  24.81952 -282.57748  52.47370
Ridge minimum MSE= 16.43541 at K= 0.12
P-value for F-test ( 2.53802 , 72.86528 ) = 3.333628e-10
------------------------------------------------------------------
```

0.95-CI beta3 (-0.68, -0.01)
0.95-CI beta4 (-0.36, 0.37)

PURDUE UNIVERSITY.

Research Question 4

Does complete series have the same impact as the booster shot with the first dose and death cases in the model?

- F test

- Do not have the same effect

```
> # F test
> anova(covidReduced.mod,covidTrans.mod)
Analysis of Variance Table

Model 1: newy ~ newx2 + combined + newx5
Model 2: newy ~ newx2 + newx3 + newx4 + newx5
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     72 1.7341
2     71 1.2940  1   0.44014  24.15 5.529e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test Statistic: 24.1499
Critical Value: 3.9758

Research Question 4

Does complete series have the same impact as the booster shot with the first dose and death cases in the model?

- K-fold Cross Validation

- Small RMSE

```
> # K-fold Cross validation
> set.seed(123)
> train.control <- trainControl(method = 'cv', number = 5)
> step.model1 = train(newy ~ newx2 + newx3 + newx4 + newx5, data = newdata,
method="leapBackward", tuneGrid = data.frame(nvmax = 4), trControl = train.
control)
> step.model1$results
  nvmax      RMSE  Rsquared       MAE     RMSESD  RsquaredSD
1     4 0.1352443 0.6514939 0.1084113 0.01994264  0.06654576
       MAESD
1 0.01792002
```

PURDUE
UNIVERSITY.

Research Question 5

Does the number of vaccines have the same impact on the number of death cases before and after the mask mandate was lifted in Indiana?

X_1: The number of vaccines

X_2: Mask mandate

When $X_2 = 0$, $E(Y) = \beta_0 + \beta_1 X_1$; When $X_2 = 1$, $E(Y) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X_1$

Y: The number of death cases

H_0: $\beta_3 = 0$ H_a: $\beta_3 \neq 0$

Reduced Model: $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

Full Model: $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$

PURDUE
UNIVERSITY®

Constant Variance

studentized Breusch-Pagan test

data: model
BP = 19.313, df = 3, p-value = 0.0002355

Normality

Shapiro-Wilk normality test

data: residuals(model)
W = 0.89634, p-value = 1.89e-07

Non-constant variance & violation of normality

Log Transformation On X   **Box-cox Transformation on Y**

Rechecking constant variance & normality
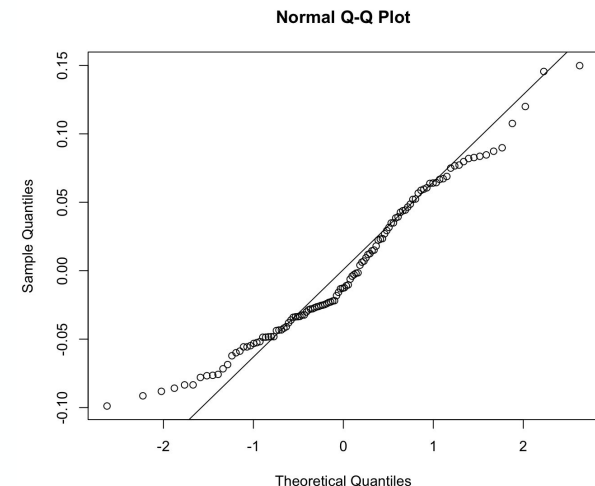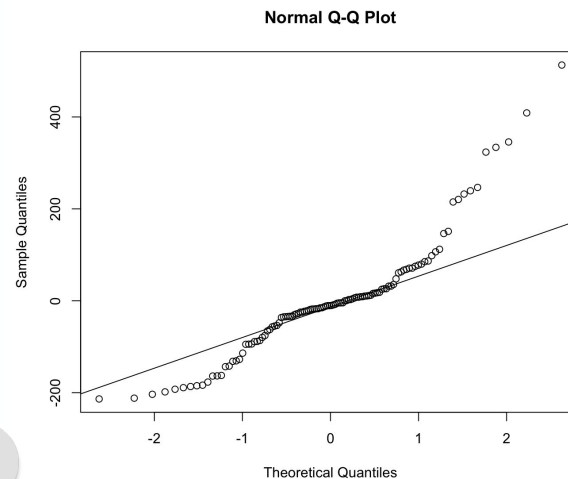
studentized Breusch-Pagan test

data: newModel
BP = 16.304, df = 3, p-value = 0.0009825

Shapiro-Wilk normality test

data: residuals(newModel)
W = 0.88327, p-value = 4.516e-08

There is an improvement on non-constant variance and normality

Research Question 5

Does the number of vaccines have the same impact on the number of death cases before and after the mask mandate was lifted in Indiana?

X_1: The number of vaccines

X_2: Mask mandate

When $X_2 = 0$, $E(Y) = \beta_0 + \beta_1 X_1$; When $X_2 = 1$, $E(Y) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X_1$

Y: The number of death cases

$H_0: \beta_3 = 0$ $H_a: \beta_3 \neq 0$

Reduced Model: $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2$

Full Model: $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$

According to OLS model, the interaction factor between two parameters is **insignificant**.

```
> anova(newModel)
Analysis of Variance Table

Response: death
                    Df  Sum Sq  Mean Sq   F value    Pr(>F)
TotalVac             1 0.09824 0.098235  31.1020  1.718e-07 ***
factor(Mask)         1 0.07482 0.074823  23.6896  3.737e-06 ***
TotalVac:factor(Mask) 1 0.00057 0.000566   0.1793     0.6728
Residuals          112 0.35375 0.003158
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
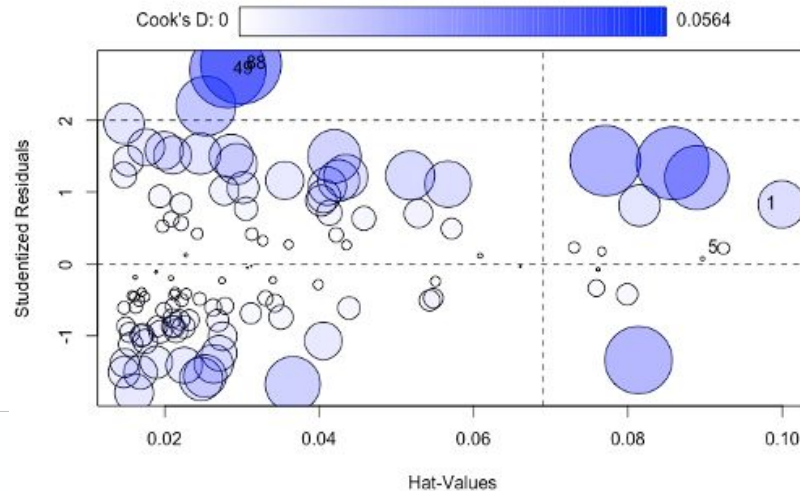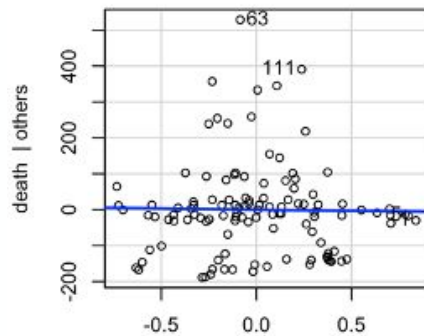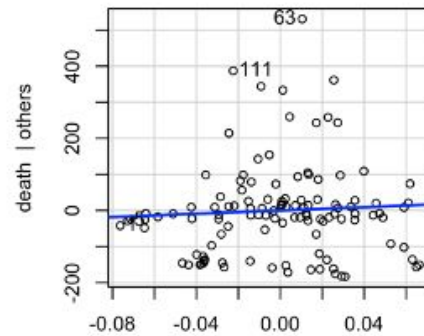
PURDUE UNIVERSITY.

Added-Variable Plots
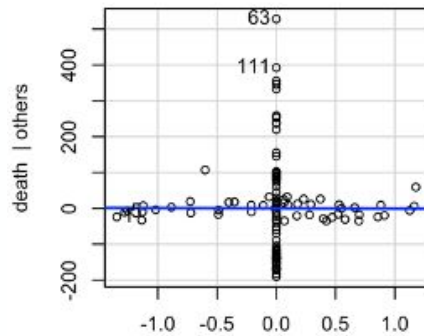
Marginal Effect

Total number of vaccines and interaction factor **don't provide much additional information**.

X and Y Outliers

Both X and Y have no obvious outliers.

Multicollinearity

TotalVac	factor(Mask)	TotalVac:factor(Mask)
5.803335	224.744249	268.974980

Have multicollinearity issue, but it's expected

Improve multicollinearity by Ridge Regression

```
> vif(RModel) # 0.06
         TotalVac factor(Mask)1 TotalVac:factor(Mask)1
k=0       5.80334     224.74425              268.97498
k=0.02    2.11490       2.56865                2.66031
k=0.04    1.87480       1.14226                0.99427
k=0.06    1.68544       0.80927                0.63227
k=0.08    1.52610       0.66171                0.48709
k=0.1     1.38974       0.57399                0.40919
k=0.12    1.27187       0.51299                0.35968

> summary(RModel)

Call:
lmridge.default(formula = death ~ TotalVac + TotalVac * factor(Mask),
    data = data, K = 0.06)


Coefficients: for Ridge parameter K= 0.06
                         Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept                  0.3492        1.8637      0.8494       2.1943  0.0303 *
TotalVac                  -0.0009       -0.0104      0.0724      -0.1441  0.8857
factor(Mask)1             -0.0379       -0.2010      0.0502      -4.0081  0.0001 ***
TotalVac:factor(Mask)1    -0.0032       -0.1961      0.0443      -4.4234  <2e-16 ***
---
```

From Ridge Regression model, we find that the parameter: total number of vaccines doesn't play an significant role in the model.

PURDUE
UNIVERSITY®

Bootstrapping

The 95% CI of X1 is **[-0.0136, 0.0109]**
The 95% CI of X2 is [-0.0602, -0.0160]
The 95% CI of the interaction factor between X1 and X2 is [-0.0042, -0.0019].

K-Fold Validation

```
> step.model1$results
  nvmax       RMSE  Rsquared        MAE     RMSESD RsquaredSD      MAESD
1     4 0.05637614 0.3503281 0.04750515 0.008679268  0.1750909 0.006109299
```

RMSE=0.056376 means **the model can make a good prediction.**

Conclusion:

1. The number of death cases were lower before the mask mandate was lifted.
2. Total number of vaccines isn't significant when it's in the same model with the categorial parameter mask mandate.

PURDUE UNIVERSITY.

Conclusion

Nature of Study

In this project, we use linear regression models to determine how factors including the first dose of vaccine, the complete series of vaccines (second dose), the booster, the total number of vaccines people take, and the categorical factor like whether the mask mandate was lifted in Indiana affects the number of infections and the number of deaths.

Our study is a **confirmatory observational study**

Real-world effectiveness of COVID-19 vaccines: a literature review and meta-analysis

Caifang Zheng [1] • Weihao Shao [1] • Xiaorui Chen • Bowen Zhang • Gaili Wang • Weidong Zhang 👤 ✉ •

Show footnotes

Open Access • Published: November 16, 2021 • DOI: https://doi.org/10.1016/j.ijid.2021.11.009 •

Check for updates

An evidence review of face masks against COVID-19

Jeremy Howard 🆔 ✉, Austin Huang, Zhiyuan Li 🆔, +15, and Anne W. Rimoin 🆔 Authors Info & Affiliations

PURDUE UNIVERSITY.

Conclusion

Expectations

1. A reduction in the number of infections after people take the first dose of COVID-19 vaccine, and continuous reduction when people take the second dose.
2. Booster can further reduce the number of infections in Indiana.
3. See the same effect of vaccines before and after the mask mandate was lifted which can further confirm that vaccines do have a strong effectiveness in preventing COVID-19 even when people didn't wear a mask.
4. Through wearing masks and getting vaccinated, we also expected to see a much lower number of death cases.

All our models have very strong predictive ability, but not all our expectations are reached.

PURDUE UNIVERSITY.

Conclusion

Pros & Cons

Pros

1. The database of our project is **up-to-date** since the pandemic just happened in the last three years.
2. The project can be meaningful since the pandemic is not completely over.
3. We learn how to deal with real-world data which is much problematic than database we used in homeworks.

Cons

1. **The scale of research questions can't be handled by relatively small database.**
2. Not all violations of data such as non-constant variance and normality were solved in our model.
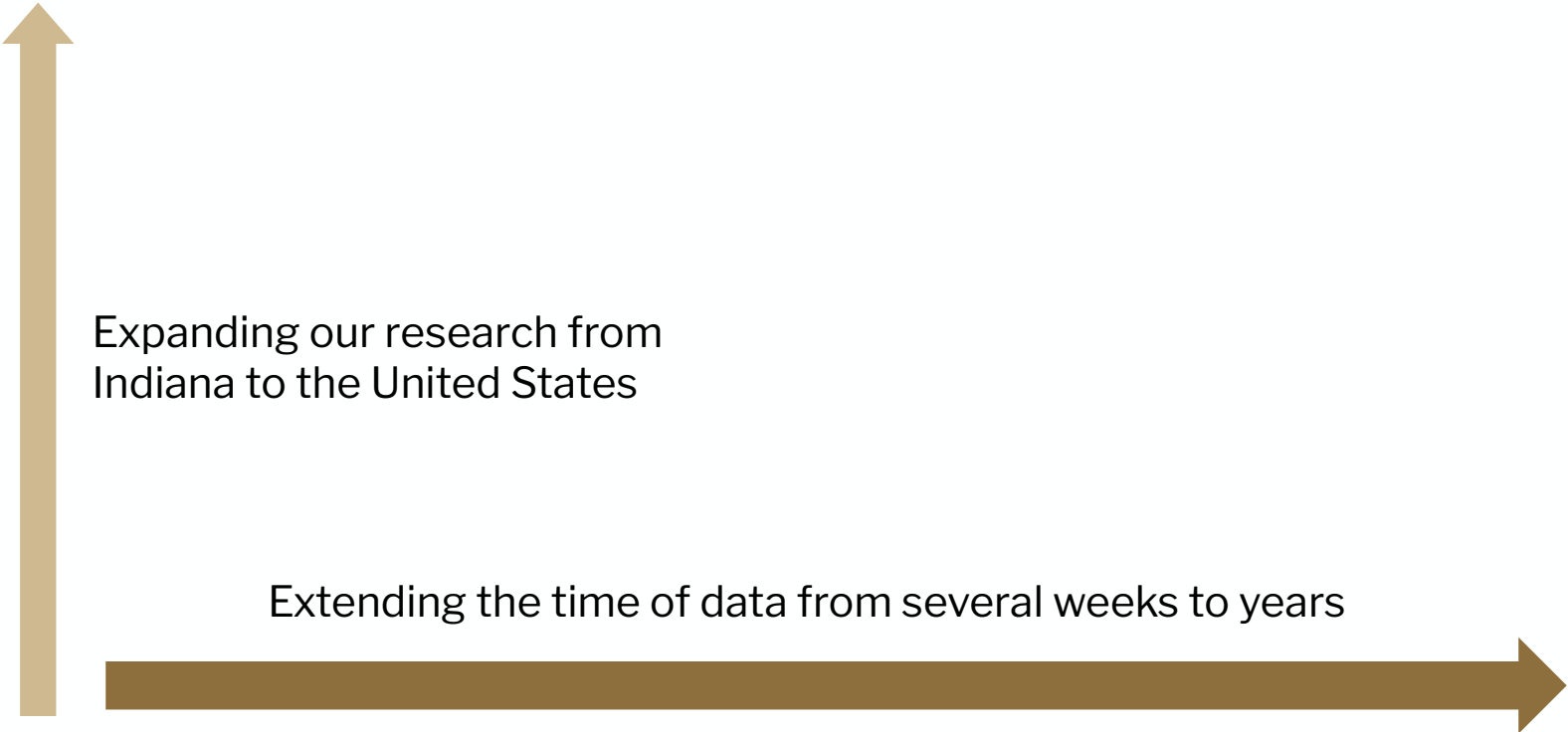3. Project wasn't planned well. We didn't set aside more time for emergency.

Conclusion

Next step

The scale of research questions can't be handled by relatively small database

Scale up current database

Expanding our research from
Indiana to the United States

Extending the time of data from several weeks to years

PURDUE
UNIVERSITY.

Thank You