Vishnu Vardhan 2020480
Vedant Bothra 2020260
Shivam Yadav 2020471
Shelja Agarwal 2020470
Saanvi Sehwag 2020462

Variable v41 - Percentage of infant deaths due to Asphyxia (to total reported infant deaths)

Q1)
a)

```
Call:
lm(formula = d$v41 ~ ., data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-16.287  -7.890  -2.585   4.624  92.447

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.410e+00  2.083e-01  35.567  < 2e-16 ***
gdp          5.888e-06  4.406e-07  13.364  < 2e-16 ***
beds        -2.219e-06  1.779e-06  -1.247  0.21238
tap         -1.135e-02  4.004e-03  -2.836  0.00457 **
v28          7.737e-02  1.043e-02   7.418 1.25e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.41 on 15483 degrees of freedom
  (10250 observations deleted due to missingness)
Multiple R-squared:  0.03605,   Adjusted R-squared:  0.0358
F-statistic: 144.8 on 4 and 15483 DF,  p-value: < 2.2e-16
```

v28 = Percentage of newborns having a weight less than 2.5 kg (to newborns weighed at birth) for Rabi season

```
Call:
lm(formula = d$v41 ~ ., data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-15.409  -7.930  -2.771   4.633  92.199

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.678e+00  1.982e-01  38.732  < 2e-16 ***
gdp          5.136e-06  4.055e-07  12.665  < 2e-16 ***
beds        -1.782e-08  1.692e-06  -0.011  0.99160
tap         -1.031e-02  3.828e-03  -2.693  0.00709 **
v28          6.089e-02  9.513e-03   6.401 1.58e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.51 on 17326 degrees of freedom
  (11498 observations deleted due to missingness)
Multiple R-squared:  0.03269,   Adjusted R-squared:  0.03247
F-statistic: 146.4 on 4 and 17326 DF,  p-value: < 2.2e-16
```

v28 = Percentage of newborns having a weight less than 2.5 kg (to newborns weighed at birth) for Kharif season

Considering v16 = Percentage of safe deliveries (to total reported deliveries) Rabi season

```
Call:
lm(formula = d$v41 ~ ., data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-15.069  -7.899  -2.479   4.472  92.153

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.434e+00  6.888e-01  -6.437 1.25e-10 ***
gdp          4.095e-06  4.054e-07  10.101  < 2e-16 ***
beds        -2.479e-07  1.676e-06  -0.148    0.882
tap         -3.553e-02  3.885e-03  -9.146  < 2e-16 ***
v16          1.544e-01  8.034e-03  19.220  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.41 on 17326 degrees of freedom
  (11498 observations deleted due to missingness)
Multiple R-squared:  0.05065,   Adjusted R-squared:  0.05043
F-statistic: 231.1 on 4 and 17326 DF,  p-value: < 2.2e-16
```

Considering v16 = Percentage of safe deliveries (to total reported deliveries) Kharif season

Considering both v28 and v16 in the model. For Rabi

Considering both v16 and v28 for Kharif

For Rabi season

```
        Pearson's product-moment correlation

data:  d$v41 and d$v28
t = 10.538, df = 15710, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.06823566 0.09928911
sample estimates:
       cor
0.08378273
```

v41 and v28 has the positive correlation

```
        Pearson's product-moment correlation

data:  d$v41 and d$v16
t = 24.128, df = 15710, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1739107 0.2040664
sample estimates:
      cor
0.1890331
```

v41 and v16 has the positive correlation

For Kharif season

```
        Pearson's product-moment correlation

data:  d$v41 and d$v16
t = 23.974, df = 17586, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1635512 0.1921740
sample estimates:
      cor
0.1779002
```

v41 and v16 has the positive correlation

```
        Pearson's product-moment correlation

data:  d$v41 and d$v28
t = 9.6621, df = 17586, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.05795015 0.08735212
sample estimates:
       cor
0.07266692
```

v41 and v28 has the positive correlation

```
Call:
lm(formula = d$v41 ~ ., data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-15.083  -8.028  -2.417   4.473  92.863

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.549e+00  7.176e-01  -6.339 2.37e-10 ***
gdp          3.575e-06  4.134e-07   8.649  < 2e-16 ***
beds         2.003e-07  1.726e-06   0.116  0.90762
tap         -3.626e-02  4.030e-03  -8.998  < 2e-16 ***
v16          1.559e-01  8.358e-03  18.654  < 2e-16 ***
v28          2.922e-02  9.677e-03   3.020  0.00253 **
index       -2.294e-02  1.152e-02  -1.992  0.04641 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.48 on 16850 degrees of freedom
Multiple R-squared:  0.04752,   Adjusted R-squared:  0.04718
F-statistic: 140.1 on 6 and 16850 DF,  p-value: < 2.2e-16
```

Considering the gdp, beds, tap, v16, v28, index in Kharif

```
Call:
lm(formula = d$v41 ~ ., data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-15.785  -8.029  -2.189   4.421  92.439

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.803e+00  7.012e-01  -6.849 7.71e-12 ***
gdp          4.264e-06  4.462e-07   9.554  < 2e-16 ***
beds        -1.999e-06  1.770e-06  -1.129   0.2590
tap         -3.760e-02  4.215e-03  -8.920  < 2e-16 ***
v16          1.547e-01  8.292e-03  18.658  < 2e-16 ***
v28          4.593e-02  1.056e-02   4.351 1.37e-05 ***
index        2.160e-02  1.101e-02   1.962   0.0498 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.35 on 15176 degrees of freedom
  (10555 observations deleted due to missingness)
Multiple R-squared:  0.0542,    Adjusted R-squared:  0.05383
F-statistic: 144.9 on 6 and 15176 DF,  p-value: < 2.2e-16
```

Considering the gdp, beds, tap, v16, v28, index in Rabi

Considering R value it is clearly evident that considering both v28, v16, index would be better model.
v28 Percentage of newborns having weight less than 2.5 kg (to newborns weighed at birth)
v16 Percentage of safe deliveries (to total reported deliveries)
gdp It is the gdp of that state on that particular year
beds It is the statewise number of beds as of 2020
tap It is district wise tap water access

v41 Percentage of infant deaths due to Asphyxia (to total reported infant deaths)

Birth asphyxia is nowadays one of the leading causes of death of babies especially in low and middle income countries. While there are other causes of asphyxia like caesarean and assisted deliveries, low birth weight is a major cause or determinant of asphyxia.

The weight of the baby at the time of delivery or birth is considered normal if it is somewhere between 2500gm to 4000gm, but it is considered as low if under 2500gm and the baby has greater chances or risks of asphyxia.
According to many research and surveys, around 10.3 percent of the total cases of asphyxia have low birth weight as a major determinant. And babies born with low birth weight are 3.74 times more likely to develop birth asphyxia compared to normal birth weight. So low birth weight has 4-fold higher odds of birth asphyxia compared to normal birth weight.
Low birth weight is generally associated with maternal medical conditions such as hypertension and diabetes that increase the burden of birth asphyxia.

Yes, safe baby deliveries relate to asphyxia. It is like the baby may be born safe but may suffer from asphyxia just after that. That's called birth asphyxia. Primarily there is a lack of oxygen and blood flow to the brain.
In simple terms, a baby's organs do not get enough oxygen and nutrients. It may be before, during, or right after the birth.
The amount of harm to the baby depends on various factors like how long the baby does not get oxygen, how low the level of oxygen is and etc.
Usually, babies with mild or moderate asphyxia may recover fully but if cells do not get enough oxygen for a longer period of time, then a baby may have a permanent injury.
There is a robust relationship between birth asphyxia and per capita GDP in various states of India. On average, a 1 percent decrease in per capita GDP results in an increase of between 0.34 to 0.40 per 1000 children born with birth asphyxia, it is true with infant mortality in general.
It has been observed that states with better income and GDP have relatively low cases of birth asphyxia, and infant mortality.
Other resources also suggest that 10 percent increase ini GDP would decrease birth asphyxia by 10 percent.

Tap water facility effects the asphixiya as it is an inexpensive cooling technique. It is an in expensive equipment to administer therapeutic hypothermia. Therapeutic Hypothermia a systematic review of 11 randomized controlled trials (1505 term and late preterm infants with moderate/severe encephalopathy and evidence of intrapartum asphyxia) was conducted in 2013 by the Cochrane Collaboration. It

concluded that, TH resulted in a statistically significant and clinically important reduction in the combined outcome of mortality or major neurodevelopmental disability at 18 mo of age
So Tap is included in the model.

Now, It is clearly evident that we have to consider the model with v28 and v16. The correlation coefficients for every season is positive this strengthens our assumption.

Perfect model after considering all model specification:

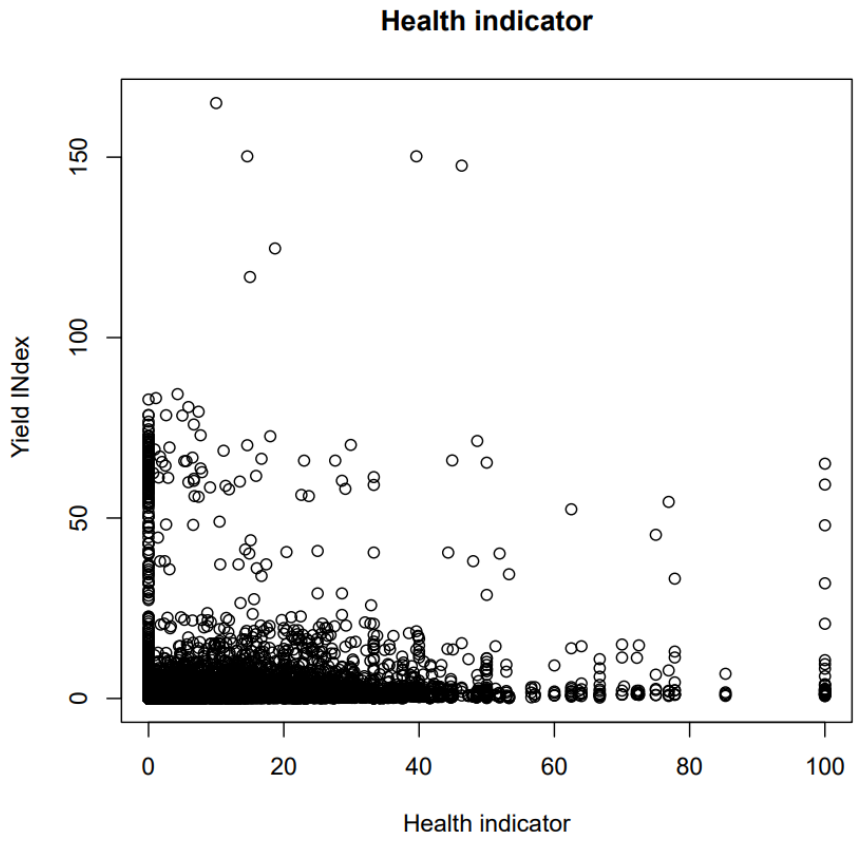
v41 ~ gdp+log(beds)+tap+index+v28
For Rabi

```
Call:
lm(formula = d$v41 ~ ., data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-16.008  -8.023  -2.543   4.651  92.226

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.729e+00  1.276e+00   6.055 1.44e-09 ***
gdp          5.291e-08  4.250e-09  12.451  < 2e-16 ***
l_beds      -1.378e-02  1.318e-01  -0.105   0.9167
tap         -1.107e-02  3.904e-03  -2.834   0.0046 **
v28          6.972e-02  1.073e-02   6.499 8.33e-11 ***
index        2.094e-02  1.113e-02   1.880   0.0601 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.48 on 15171 degrees of freedom
  (10561 observations deleted due to missingness)
Multiple R-squared:  0.03247,   Adjusted R-squared:  0.03216
F-statistic: 101.8 on 5 and 15171 DF,  p-value: < 2.2e-16
```

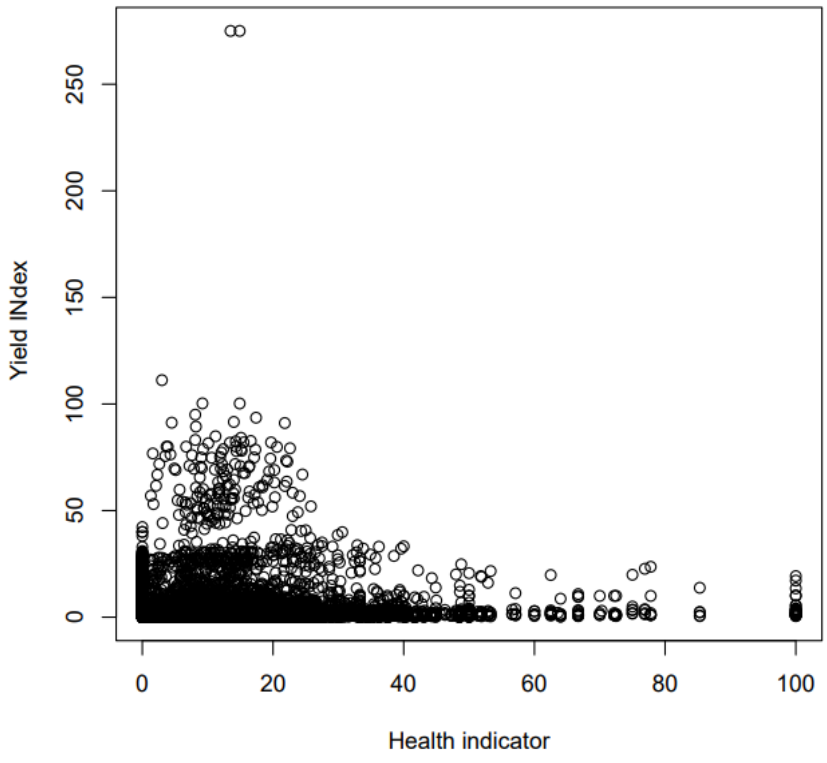
For Kharif

```
Call:
lm(formula = d$v41 ~ ., data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-15.174  -8.260  -2.722   4.611  92.893

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.051e+00  1.187e+00   6.782 1.23e-11 ***
gdp          4.847e-08  3.862e-09  12.549  < 2e-16 ***
l_beds       2.114e-02  1.227e-01   0.172 0.863268
tap         -1.368e-02  3.729e-03  -3.667 0.000246 ***
v28          4.778e-02  9.891e-03   4.831 1.37e-06 ***
index       -3.802e-02  1.140e-02  -3.337 0.000850 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.6 on 16845 degrees of freedom
  (11978 observations deleted due to missingness)
Multiple R-squared:  0.02786,   Adjusted R-squared:  0.02757
F-statistic: 96.56 on 5 and 16845 DF,  p-value: < 2.2e-16
```

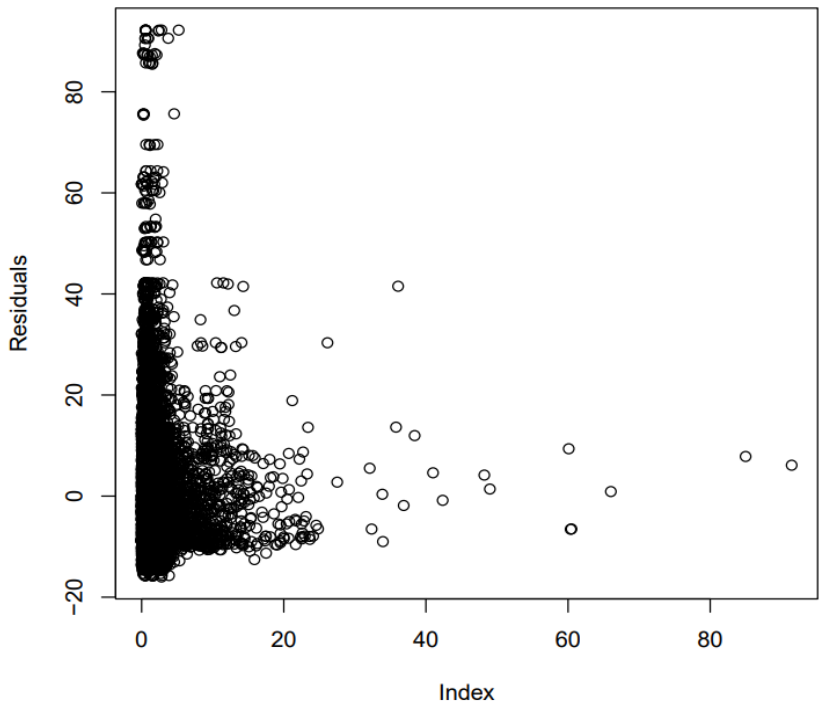
a) Kharif



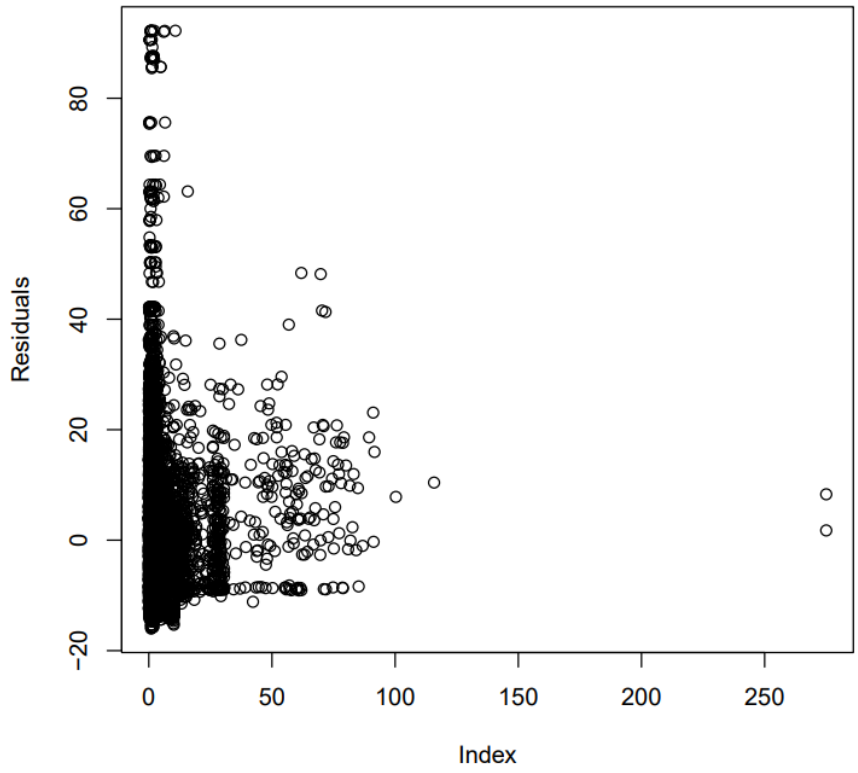
Rabi



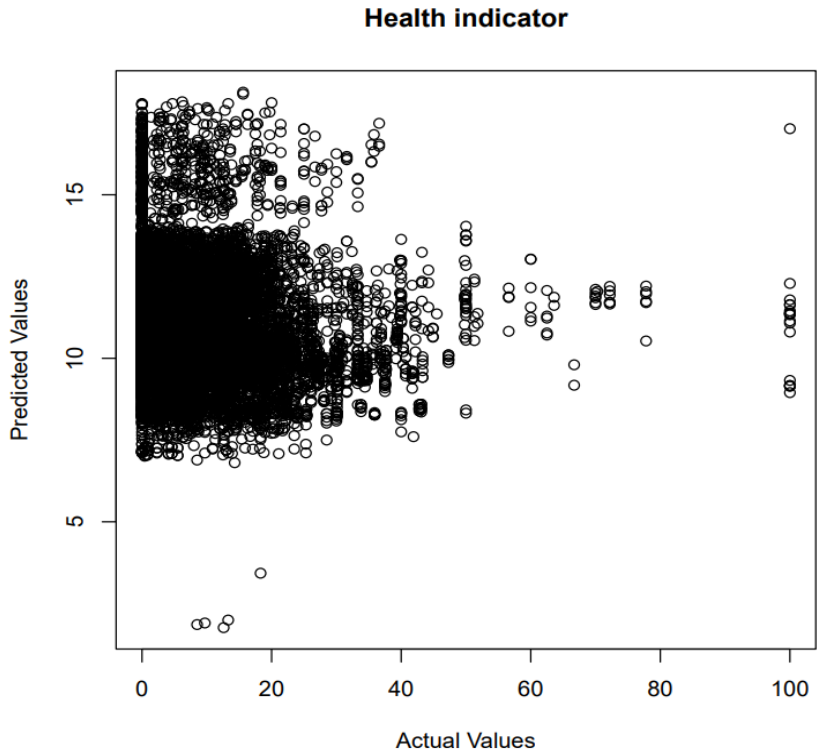
b) Kharif



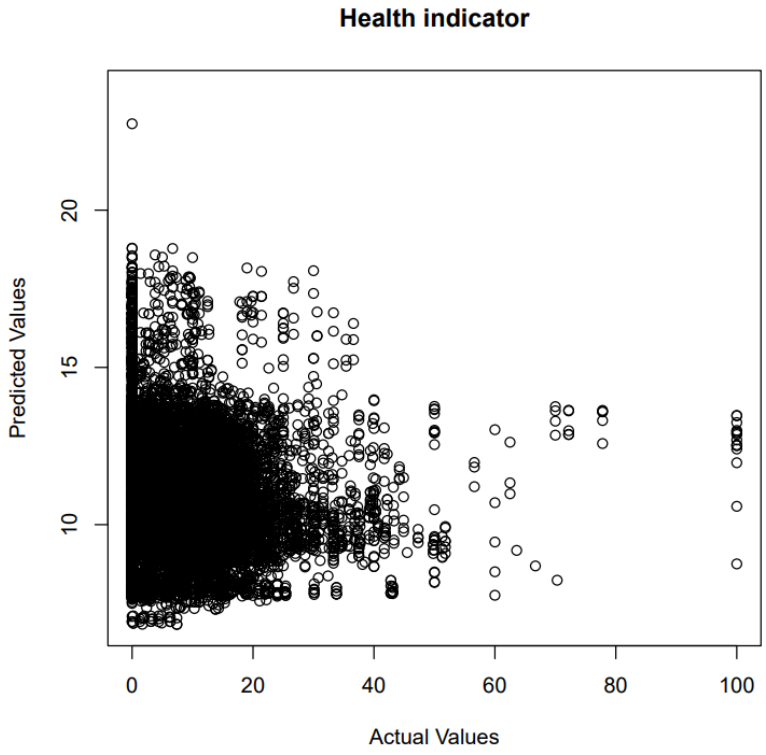
Rabi



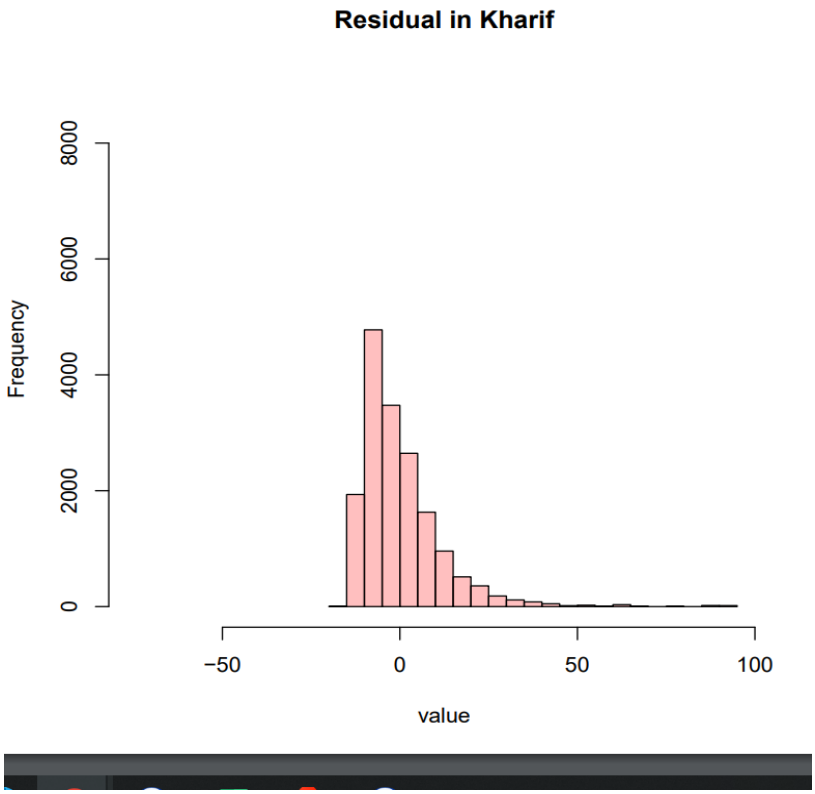
c) Kharif



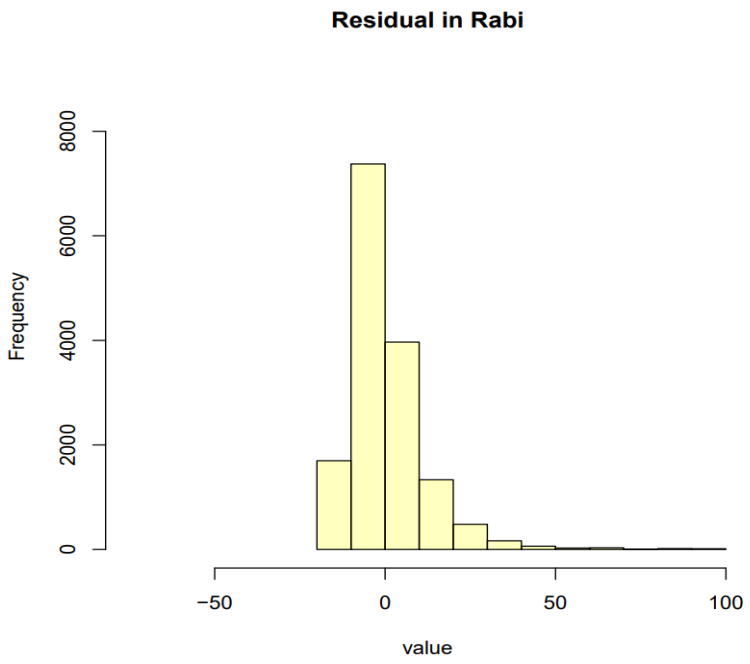
Rabi



C)
Sum of residuals for Kharif
3.484857e-11
Also observing the graph it is clear that there is very slight difference between positive values and negative values



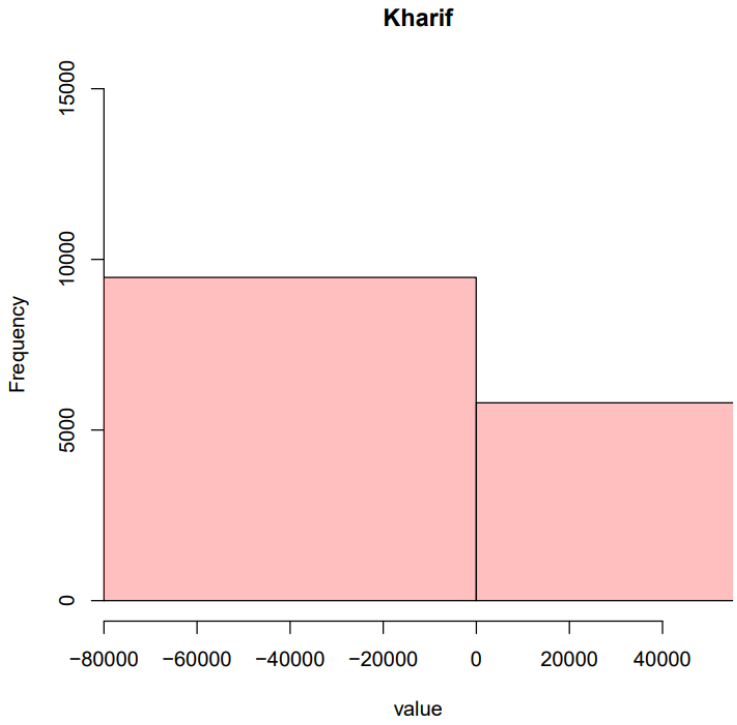
Sum of residuals for Rabi
-2.994383e-11
Also obersiving the graph it is clear that there is very slight differnce between positive values and negative values



D) <u>For kharif</u>

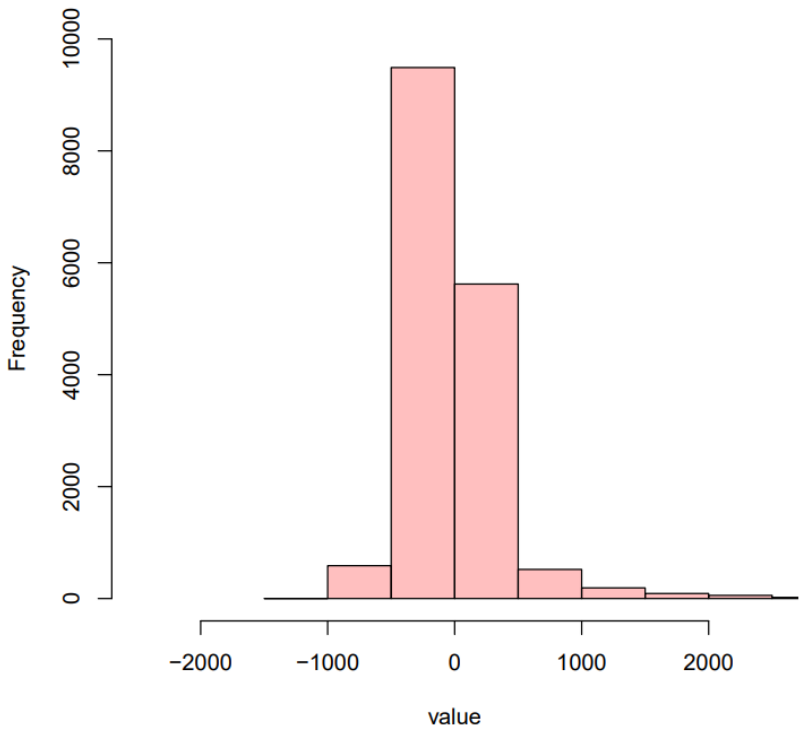
Sum of residuals* variables of GDP

317600354526

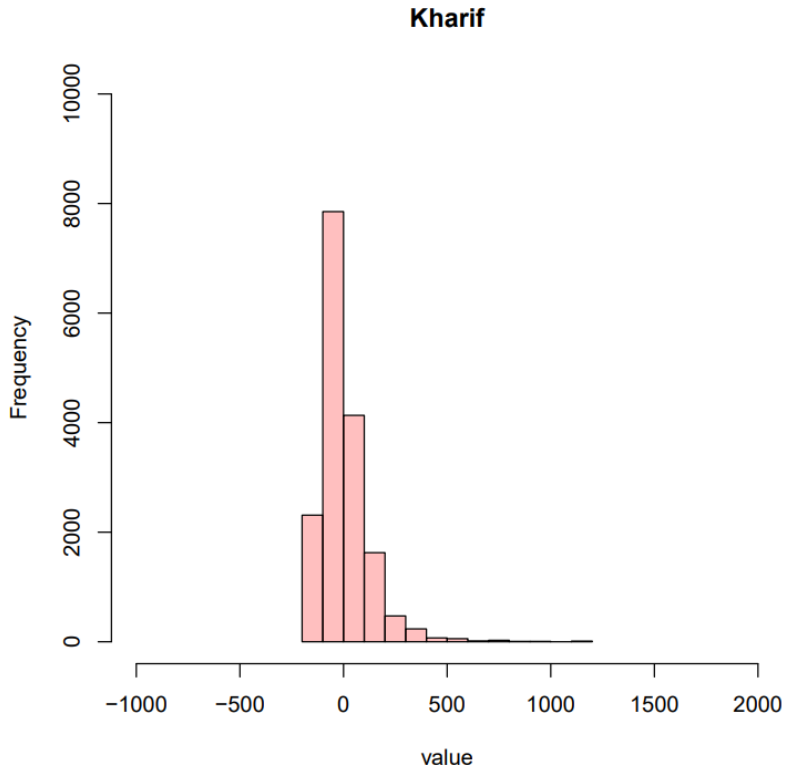


Sum of residuals* variables of tap

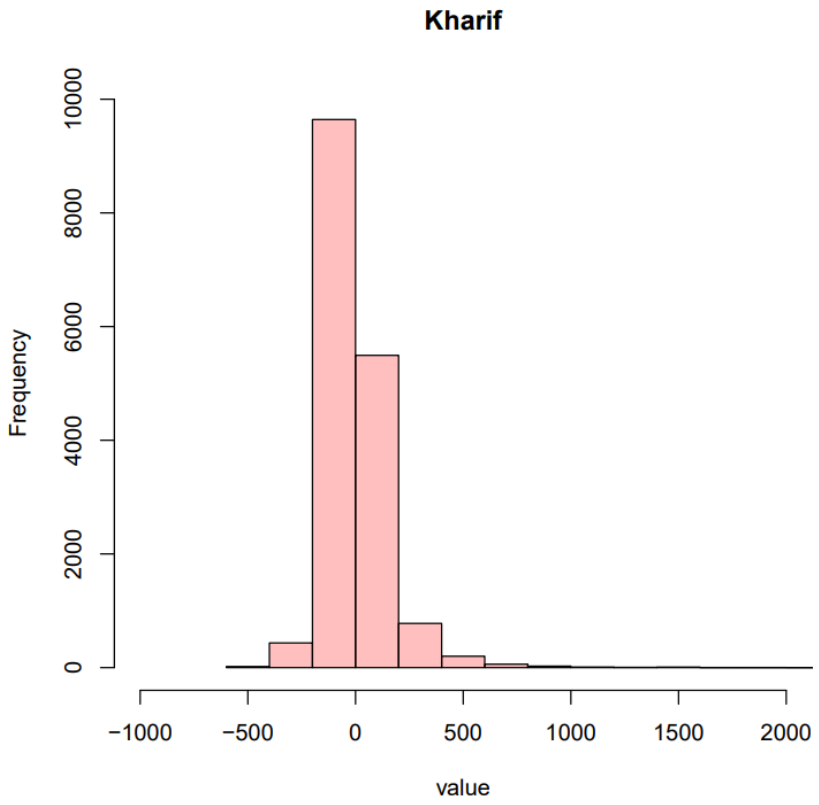
209198



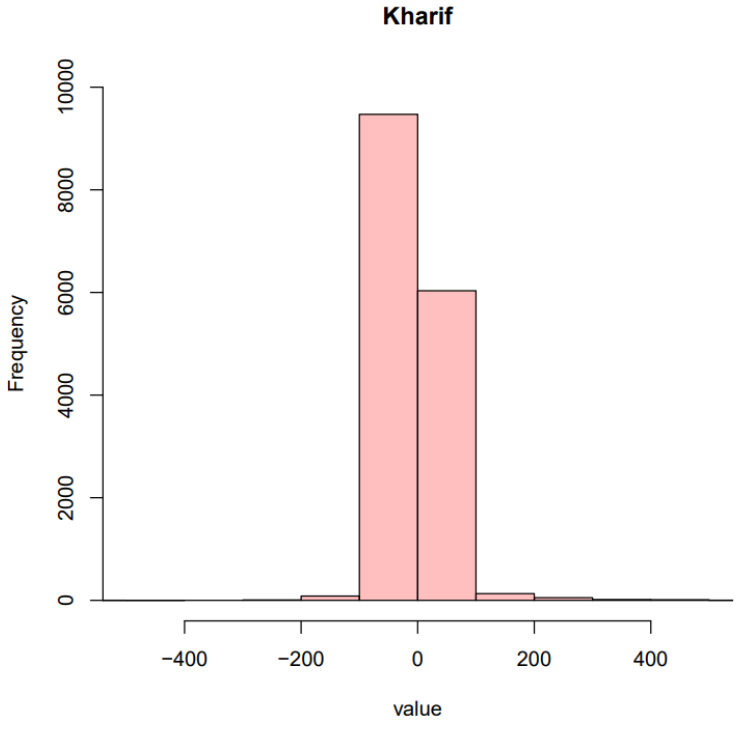
Sum of residuals* variables of l_beds
12269.53



Sum of residuals* variables of v28
8257.101



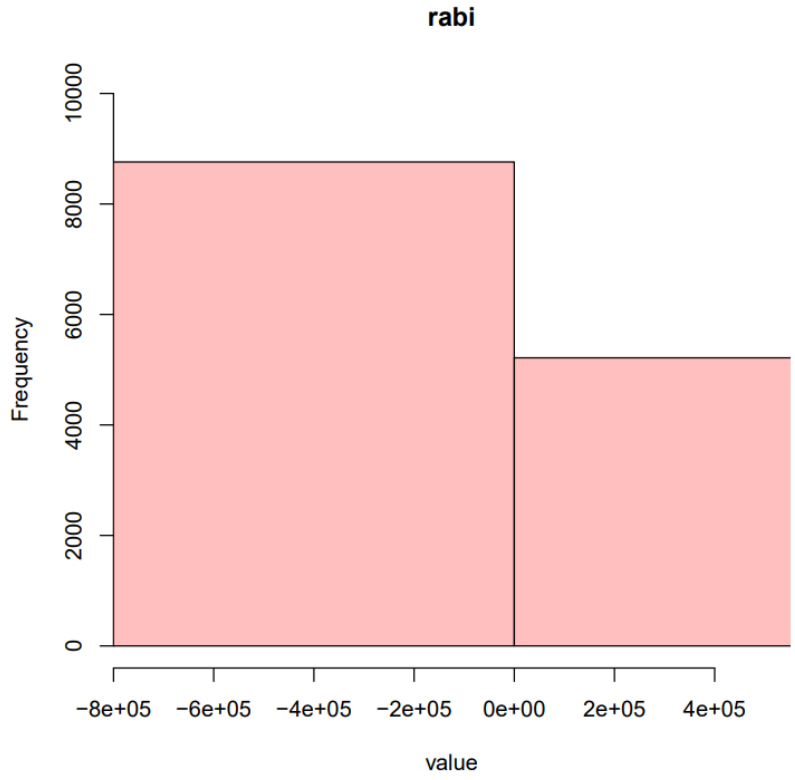
Sum of residuals* variables of index
15280.91



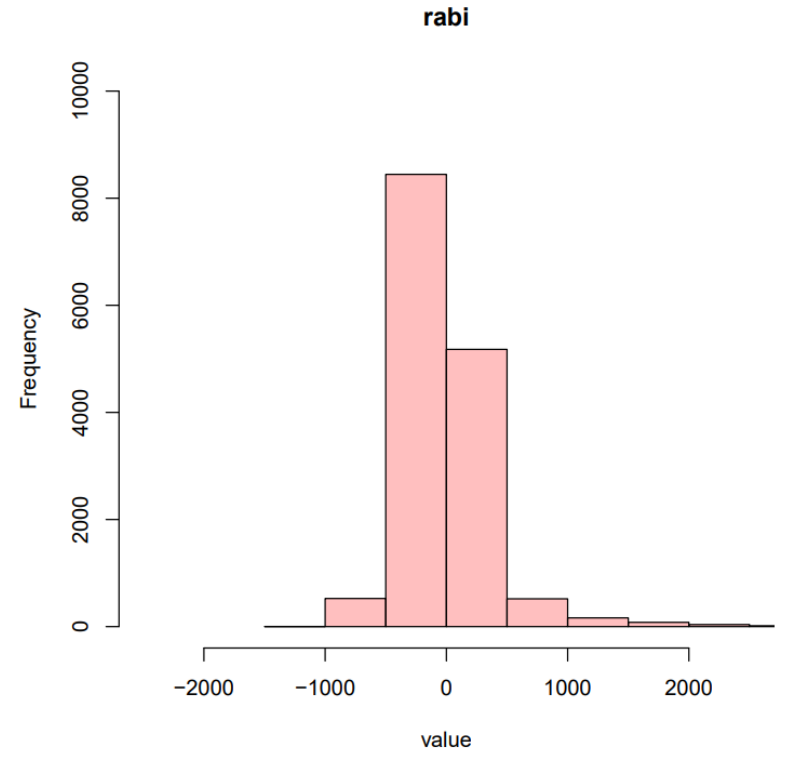
For Rabi

Sum of residuals* variables of GDP

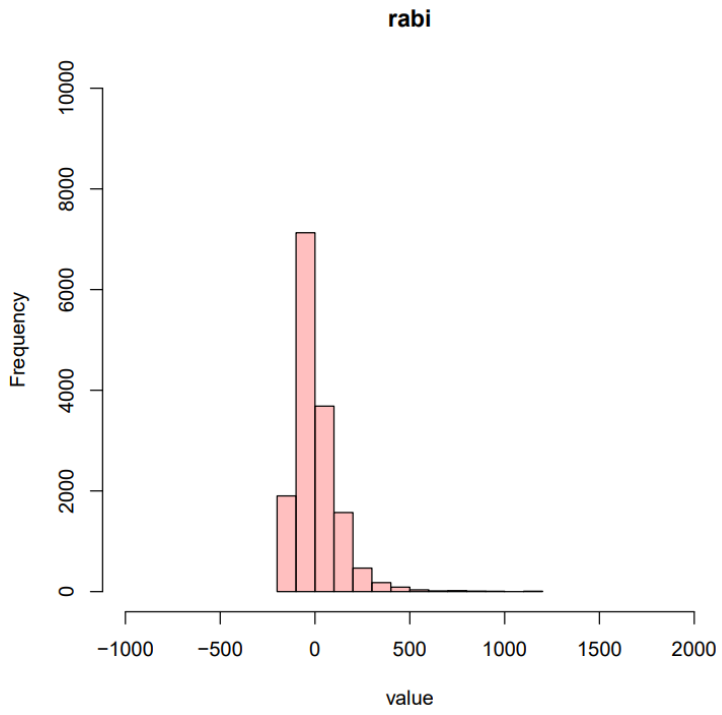
1.174374e+12



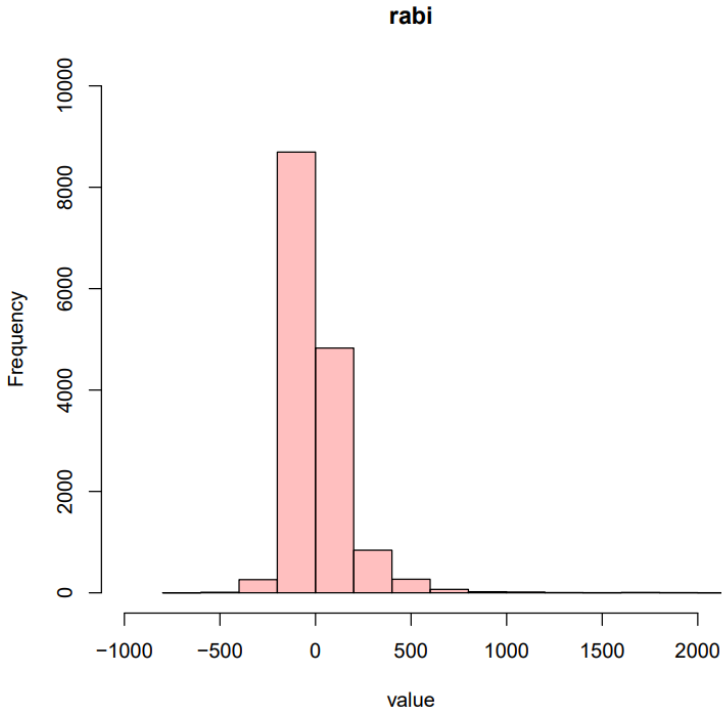
Sum of residuals* variables of tap
323607.4



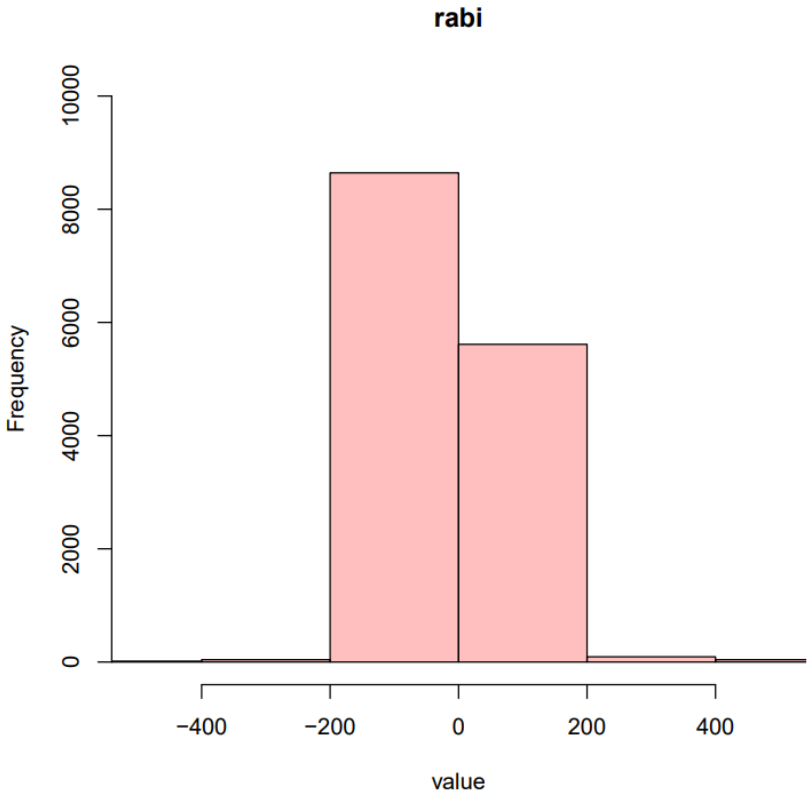
Sum of residuals* variables of l_beds
64486.85



Sum of residuals* variables of v28
38443.12



Sum of residuals* variables of index
84297.43


rabi
Frequency
0
2000
4000
6000
8000
10000
−400
−200
0
200
400
value

Q-2

a)
Assigned Health indicator is represented by: v41

Two models were made, one for kharif season and other one for the rabi season.

1)
model_kharif = lm(main.v41 ~ main.index, data = main_kharif_df)

```
> summary(model_kharif)

Call:
lm(formula = main.v41 ~ main.index, data = main_kharif_df)

Residuals:
    Min      1Q  Median      3Q     Max
-11.081  -8.881  -2.766   4.959  90.240

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.08073    0.09535 116.215   <2e-16 ***
main.index  -0.02030    0.01146  -1.771   0.0766 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.78 on 16990 degrees of freedom
  (11837 observations deleted due to missingness)
Multiple R-squared:  0.0001845, Adjusted R-squared:  0.0001257
F-statistic: 3.136 on 1 and 16990 DF,  p-value: 0.0766
```

Slope: -0.02030
Intercept: 11.08073

2)

model_rabi = lm(main.v41 ~ main.index, data = main_rabi_df)

```
> summary(model_rabi)

Call:
lm(formula = main.v41 ~ main.index, data = main_rabi_df)

Residuals:
    Min      1Q  Median      3Q     Max
-12.024  -8.913  -2.672   5.021  89.122

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.86414    0.10153 107.009   <2e-16 ***
main.index   0.02745    0.01117   2.458    0.014 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.69 on 15325 degrees of freedom
  (10411 observations deleted due to missingness)
Multiple R-squared:  0.000394,  Adjusted R-squared:  0.0003288
F-statistic: 6.041 on 1 and 15325 DF,  p-value: 0.01399
```

Slope:0.02745
Intercept:10.86414

b)

Now, these estimates for both the seasons are treated as true population parameters, and the Monte Carlo Simulations are carried out.

n = sample size(80% of the dataset chosen randomly)

N= number of iterations = 2000.

1) Kharif:

```
> stargazer(monte_estimates_kharif_vec,type="text")

================================================
Statistic   N    Mean   St. Dev.   Min    Max
------------------------------------------------
beta_1    2,000 -0.020   0.008    -0.048 0.007
beta_0    2,000 11.082   0.047    10.920 11.218
------------------------------------------------
```

2) Rabi:

```
================================================
Statistic   N    Mean   St. Dev.   Min    Max
------------------------------------------------
beta_1    2,000 0.027    0.004    0.014  0.043
beta_0    2,000 10.866   0.052    10.694 11.025
------------------------------------------------
```

Now if we increase the sample size for our monte carlo simulation, based on the consistency property of the ols estimator the standard deviation should become smaller, and the mean of b_0 and b_1 would get closer to the population parameters.

n = sample size(90% of the dataset chosen randomly)
N= number of iterations = 2000.

1) Kharif:

Statistic	N	Mean	St. Dev.	Min	Max
beta_1	2,000	-0.020	0.005	-0.046	-0.006
beta_0	2,000	11.081	0.031	10.955	11.189

2) Rabi:

Statistic	N	Mean	St. Dev.	Min	Max
beta_1	2,000	0.027	0.003	0.019	0.035
beta_0	2,000	10.864	0.032	10.759	10.975

As we can see from the table for both the season's, the mean has become closer to the population parameters, and the sd of the estimates have become smaller.
As the sample size becomes larger, the estimate beta_0^ and beta_1^ converges to the true value β_0 and β_1.

Q-3

a)

One model for the Kharif season and the other for the Rabi season.

model_kharif=lm(v41~log(beds)+gdp+tap+index+v28+D_south+D_north+D_central+D_west+D_east, data=main_K)
model_rabi=lm(v41~log(beds)+gdp+tap+index+v28+D_south+D_north+D_central+D_west+D_east, data=main_R)

We don't need to include a dummy variable for D_north_east because to represent 6 different groups only 5 dummy variables are required.

Model_kharif:

```
Call:
lm(formula = v41 ~ log(beds) + gdp + tap + index + v28 + D_south +
    D_north + D_central + D_west + D_east, data = main_K)

Residuals:
    Min      1Q  Median      3Q     Max
-20.346  -7.107  -2.349   4.115  92.959

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.990e+01  1.554e+00  12.807  < 2e-16 ***
log(beds)   -1.287e+00  1.748e-01  -7.361 1.91e-13 ***
gdp          9.028e-06  4.602e-07  19.617  < 2e-16 ***
tap         -6.774e-03  4.354e-03  -1.556 0.119766
index       -2.521e-03  1.115e-02  -0.226 0.821130
v28          1.676e-01  1.059e-02  15.825  < 2e-16 ***
D_south      4.338e+00  4.734e-01   9.163  < 2e-16 ***
D_north     -5.100e+00  4.124e-01 -12.367  < 2e-16 ***
D_central   -4.902e+00  3.540e-01 -13.851  < 2e-16 ***
D_west      -5.304e+00  4.496e-01 -11.797  < 2e-16 ***
D_east       1.412e+00  3.914e-01   3.609 0.000308 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.02 on 16846 degrees of freedom
  (11972 observations deleted due to missingness)
Multiple R-squared:  0.1226,    Adjusted R-squared:  0.1221
F-statistic: 235.5 on 10 and 16846 DF,  p-value: < 2.2e-16
```

Model_Rabi:

```
Call:
lm(formula = v41 ~ log(beds) + gdp + tap + index + v28 + D_south +
    D_north + D_central + D_west + D_east, data = main_R)

Residuals:
    Min      1Q  Median      3Q     Max
-20.336  -7.230  -2.297   4.146  92.660

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.003e+01  1.662e+00  12.057  < 2e-16 ***
log(beds)   -1.369e+00  1.868e-01  -7.327 2.47e-13 ***
gdp          9.484e-06  5.076e-07  18.683  < 2e-16 ***
tap         -6.921e-03  4.704e-03  -1.471   0.1413
index        2.269e-02  1.079e-02   2.104   0.0354 *
v28          1.763e-01  1.148e-02  15.358  < 2e-16 ***
D_south      4.549e+00  4.950e-01   9.190  < 2e-16 ***
D_north     -4.494e+00  4.250e-01 -10.574  < 2e-16 ***
D_central   -4.442e+00  3.755e-01 -11.830  < 2e-16 ***
D_west      -5.089e+00  4.818e-01 -10.561  < 2e-16 ***
D_east       9.441e-01  3.738e-01   2.526   0.0116 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.95 on 15172 degrees of freedom
  (10555 observations deleted due to missingness)
Multiple R-squared:  0.1195,    Adjusted R-squared:  0.119
F-statistic:   206 on 10 and 15172 DF,  p-value: < 2.2e-16
```

b)

Let the H0:
Difference between the average value of v41 between different state groups is zero, which means there is no structural break between different state groups.

Ha(Alternate Hypothesis):

Difference between the average value of v41 between different state groups is not equal to zero, which means there is a structural break between different state groups.

On conduction the t test between, different state groups we get to observe that the p value is very close to zero in almost all cases.
So we reject the null hypothesis, and accept the alternative hypothesis which means there is a structural break between different state groups.

Kharif:

```
        Welch Two Sample t-test

data:  df1$v41 by df1$zone
t = -33.036, df = 5534.7, p-value < 2.2e-16
alternative hypothesis: true difference in means between group Not South and group South is not equal to 0
95 percent confidence interval:
 -7.467041 -6.630489
sample estimates:
mean in group Not South     mean in group South
               9.354016               16.402781
```

Rabi:

```
        Welch Two Sample t-test

data:  df1$v41 by df1$zone
t = -30.565, df = 5026.6, p-value < 2.2e-16
alternative hypothesis: true difference in means between group Not South and group South is not equal to 0
95 percent confidence interval:
 -7.238306 -6.365738
sample estimates:
mean in group Not South     mean in group South
                9.40670                16.20872
```

c)
Unrestricted eqn: lm(v41~log(beds)+gdp+tap+index+v28,data=main_K)
Restricted eqn: lm(v41~log(beds)+gdp+tap+index+v28+D_south, data=main_K)

Test for structural change with respect to D_south.
Ho: D_south = 0. That signifies there is no structural change.
(Null Hypothesis)

Ha: D_south != 0 That signifies there is a structural change.
(Alternative hypothesis).

alpha = 0.01 (significance level)

F_TEST starts:(Acc to notes)

F = ((rss_restricted - rss_unrestricted)/k) / (rss_unrestricted / df_unrestrcited)

For kharif :
It comes out to be F_kharif = 416.486

For Rabi:
It comes out to be F_rabi = 371.831

We reject the Ho(Null Hypothesis) because F > critical value for both the season's rabi and kharif, and accept the alternative Hypothesis .

So, there is a structural break across southern and non southern state-groups.

Kharif Unrestricted:

```
Call:
lm(formula = v41 ~ log(beds) + gdp + tap + index + v28 + D_south,
    data = main_K)

Residuals:
    Min      1Q  Median      3Q     Max
-20.139  -7.491  -2.366   4.121  92.571

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.903e+01  1.284e+00   22.61   <2e-16 ***
log(beds)   -2.345e+00  1.353e-01  -17.33   <2e-16 ***
gdp          8.952e-06  3.893e-07   23.00   <2e-16 ***
tap         -4.275e-02  3.676e-03  -11.63   <2e-16 ***
index       -1.111e-04  1.102e-02   -0.01    0.992
v28          1.545e-01  9.986e-03   15.47   <2e-16 ***
D_south      9.147e+00  2.546e-01   35.92   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.18 on 16850 degrees of freedom
  (11972 observations deleted due to missingness)
Multiple R-squared:  0.09699,   Adjusted R-squared:  0.09667
F-statistic: 301.6 on 6 and 16850 DF,  p-value: < 2.2e-16
```

Kharif Restricted:

```
> summary(model_r)

Call:
lm(formula = v41 ~ log(beds) + gdp + tap + index + v28, data = main_K)

Residuals:
    Min      1Q  Median      3Q     Max
-15.181  -8.276  -2.717   4.614  92.884

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.140e+00  1.187e+00   6.855 7.40e-12 ***
log(beds)    1.351e-02  1.227e-01   0.110 0.912371
gdp          4.861e-06  3.862e-07  12.586  < 2e-16 ***
tap         -1.429e-02  3.724e-03  -3.836 0.000125 ***
index       -3.860e-02  1.138e-02  -3.392 0.000694 ***
v28          4.777e-02  9.893e-03   4.828 1.39e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.6 on 16851 degrees of freedom
  (11972 observations deleted due to missingness)
Multiple R-squared:  0.02784,   Adjusted R-squared:  0.02755
F-statistic: 96.52 on 5 and 16851 DF,  p-value: < 2.2e-16
```

Rabi unrestricted:

```
Call:
lm(formula = v41 ~ log(beds) + gdp + tap + index + v28 + D_south,
    data = main_R)

Residuals:
    Min      1Q  Median      3Q     Max
-20.124  -7.434  -2.344   4.137  92.010

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.762e+01  1.370e+00  20.155  < 2e-16 ***
log(beds)   -2.222e+00  1.436e-01 -15.478  < 2e-16 ***
gdp          8.918e-06  4.246e-07  21.001  < 2e-16 ***
tap         -4.378e-02  3.886e-03 -11.264  < 2e-16 ***
index        4.282e-02  1.077e-02   3.975 7.07e-05 ***
v28          1.616e-01  1.073e-02  15.068  < 2e-16 ***
D_south      8.807e+00  2.660e-01  33.109  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.08 on 15176 degrees of freedom
  (10555 observations deleted due to missingness)
Multiple R-squared:  0.09759,   Adjusted R-squared:  0.09723
```

Rabi Restricted:

```
> summary(model_r)

Call:
lm(formula = v41 ~ log(beds) + gdp + tap + index + v28, data = main_R)

Residuals:
    Min      1Q  Median      3Q     Max
-16.004  -8.039  -2.531   4.641  92.196

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.830e+00  1.277e+00   6.133 8.86e-10 ***
log(beds)   -2.265e-02  1.318e-01  -0.172  0.86357
gdp          5.305e-06  4.249e-07  12.486  < 2e-16 ***
tap         -1.176e-02  3.898e-03  -3.017  0.00256 **
index        2.068e-02  1.113e-02   1.858  0.06326 .
v28          6.973e-02  1.073e-02   6.499 8.31e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.48 on 15177 degrees of freedom
  (10555 observations deleted due to missingness)
Multiple R-squared:  0.0324,    Adjusted R-squared:  0.03208
F-statistic: 101.6 on 5 and 15177 DF,  p-value: < 2.2e-16
```