Analysis of Percentage of infant deaths due to Pneumonia

Mohit Sharma (2020086)
Yash Agrawal (2020551)
Aishwary Sharma (2020490)
Ujjwal Rastogi (2020456)
Mohit Bansal (2020526)

Introduction

Introduction

A child dies of pneumonia every 39 seconds

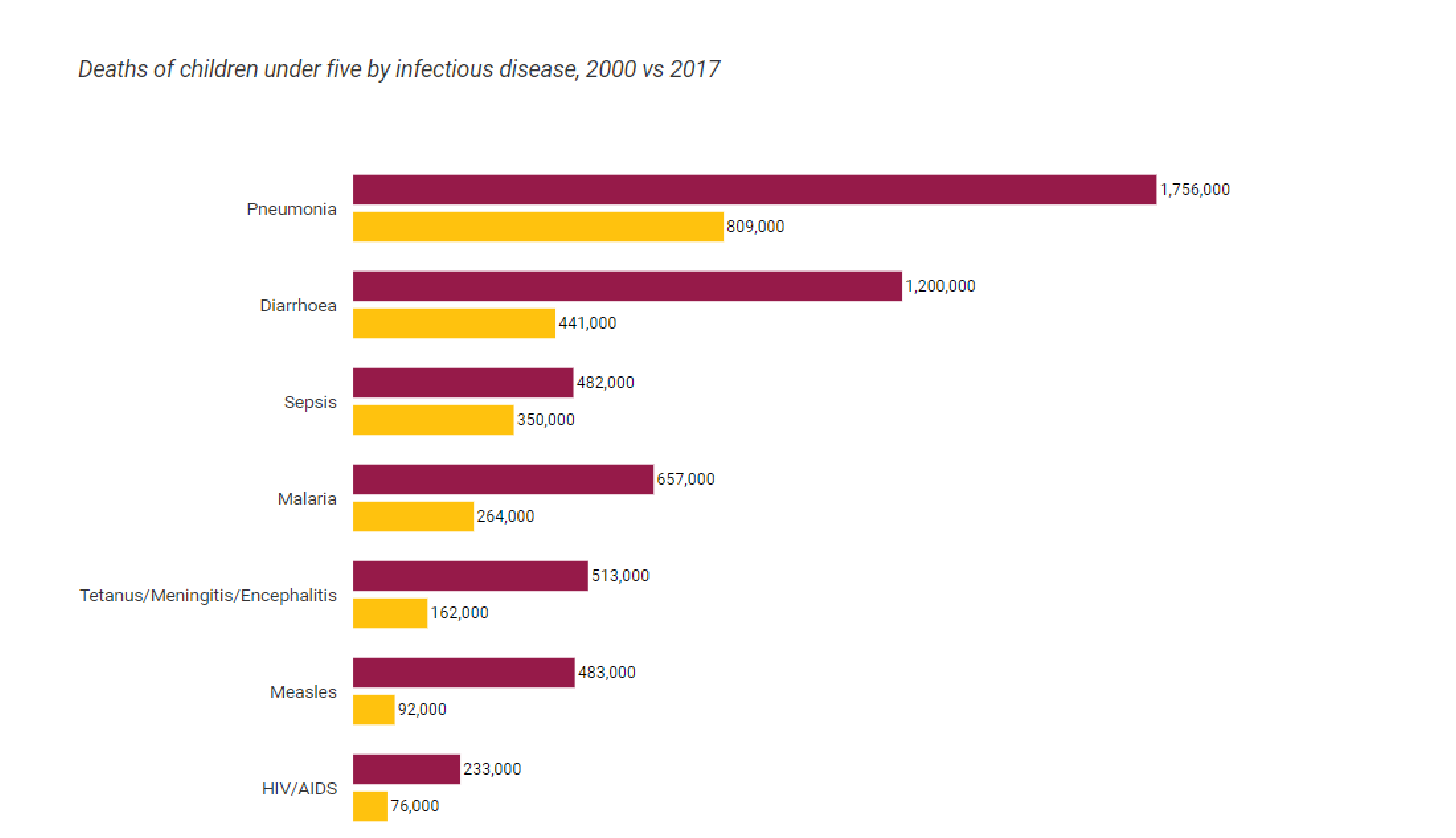
Introduction

Pneumonia is the largest infectious cause of death in children.
Globally, there is 1 case in every 71 children of Pneumonia every year.
According to 2019 Stats, 14% of the total children deaths under 5 years were due to Pneumonia.

Introduction

Deaths due to Pneumonia are high in poor regions.

Pneumonia kills more children than any other infectious disease.



Motivation

Motivation

Mortality due to childhood pneumonia is significantly connected to various poverty-related variables like malnutrition, lack of safe drinking water, indoor and outdoor air pollution, and insufficient access to health care.

Motivation

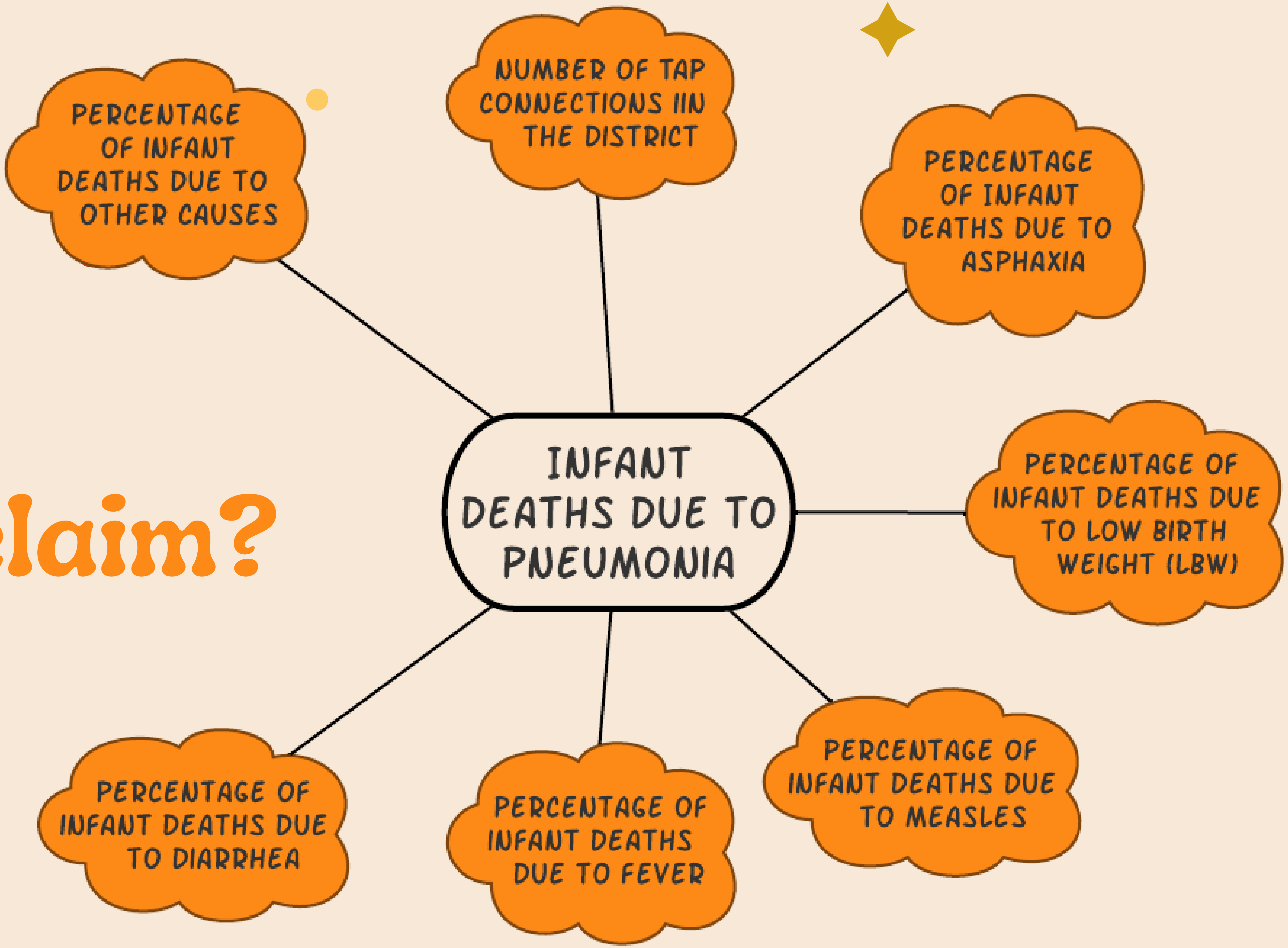
Other factors like poor health, lack of sanitation, the population of the district, and suffering from various diseases like measles, high fever, low birth weight, asphyxia, diarrhea, etc may also lead to pneumonia.

Data Used

- 29 States and 736 Districts
- Time Interval 2011 - 2019

  (Time Interval 9 years)
- Data of Rabi and Kharif Seasons
- Approximate 70,000 Data entry of more than 70 variables

What we claim?



Approach

1. We first found the possible variables that were likely to affect the cases of pneumonia.

Model 1:

$$v43_{(i,t)} = \beta_0 + \beta_1 tap_{(i,t)} + \beta_2 v41_{(i,t)} + \beta_4 v44_{(i,t)} + \beta_5 v45_{(i,t)}$$

2. Then, we added some more likely variables and included each of them in our regression model and got their coefficients value and P-value

3. We ruled out some of the variables based on their P and coefficients value.

Final Model:

$$v43_{(i,t)} = \beta_0 + \beta_1 tap_{(i,t)} + \beta_2 v41_{(i,t)} + \beta_3 v42_{(i,t)} + \beta_4 v44_{(i,t)} + \beta_5 v45_{(i,t)} + \beta_6 v46_{(i,t)} + \beta_7 v47_{(i,t)}$$

Approach

We incorporate various factors that affects the Percentage of infant deaths due to Pneumonia
We calculated the goodness of fit value for both seasons to confirm our regression analysis further.

R^2= 0.6254 (Rabi)
R^2= 0.6285 (kharif)

Variables and their Descriptions

For 35 states and union Territories over 9 Years (2011-19)

Dependent Variable

Variable Name (In the Data)	Description
v43	Percentage of infant deaths due to Pneumonia (to total reported infant deaths)

Independent Variable

Variable Name (In the Data)	Description
tap	District Wise Tap Water Access (Percentage of Households) as of 2019
v41	Percentage of infant deaths due to Asphyxia (to total reported infant deaths)
v42	Percentage of infant deaths due to Low Birth Weight (LBW) (to total reported infant deaths)
v44	Percentage of infant deaths due to Diarrhea (to total reported infant deaths)
v45	Percentage of infant deaths due to Fever (to total reported infant deaths)
v46	Percentage of infant deaths due to Measles (to total reported infant deaths)
v47	Percentage of infant deaths due to other causes (to total reported infant deaths)

Our Models

Model-1

$$v43_{(i,t)} = \beta_0 + \beta_1 tap_{(i,t)} + \beta_2 v41_{(i,t)} + \beta_3 v42_{(i,t)} + \beta_4 v44_{(i,t)} + \beta_5 v45_{(i,t)} + \beta_6 v46_{(i,t)} + \beta_7 v47_{(i,t)}$$

This model tests the relationship between the Independent variable and the Dependent variable described in the previous slide for the season Kharif.

Model-2

$$v43_{(i,t)} = \beta_0 + \beta_1 tap_{(i,t)} + \beta_2 v41_{(i,t)} + \beta_3 v42_{(i,t)} + \beta_4 v44_{(i,t)} + \beta_5 v45_{(i,t)} + \beta_6 v46_{(i,t)} + \beta_7 v47_{(i,t)}$$

This model tests the relationship between Independent variable and Dependent variable described in the previous slide for the season Rabi.

Data Summary

Data Summary Kharif Season

Acronym	Mean	Mode	Median	SD
tap	20.4785	0	8.81	24.86024
V_{41}	10.7907	0	8.1	11.70585
V_{42}	18.1456	0	16.9	13.9367
V_{44}	1.76239	0	0	6.863468
V_{45}	3.9079	0	1	9.733367
V_{46}	0.2707	0	0	3.192825
V_{47}	746.6088	100	52.35	21.71159

Data Summary Rabi Season

Acronym	Mean	Mode	Median	SD
tap	19.83234	0	7.58	24.69985
V_{41}	10.79767	0	8.2	11.6416
V_{42}	18.15516	0	17	14.01485
V_{44}	1.71813	0	0	6.746211
V_{45}	3.980289	0	1	9.81926
V_{46}	0.2024821	0	0	3.095583
V_{47}	731.2406	100	52.5	21.83915

Regression Summary of Kharif

Model 1:

$$v43_{(i,t)} = \beta_0 + \beta_1 tap_{(i,t)} + \beta_2 v41_{(i,t)} + \beta_3 v42_{(i,t)} + \beta_4 v44_{(i,t)} + \beta_5 v45_{(i,t)} + \beta_6 v46_{(i,t)} + \beta_7 v47_{(i,t)}$$

Model 1	Estimate	Standard Error	T values	P values
Intercept	63.407378	0.338712	187.201	<2e-16
tap	-0.004436	0.002130	-2.083	0.0373
v41	-0.760246	0.005958	-127.594	<2e-16
v42	-0.644428	0.004987	-129.223	<2e-16
v44	-0.578744	0.008733	-66.273	<2e-16
v45	-0.580552	0.006598	-87.987	<2e-16
v46	-0.627616	0.016391	-38.290	<2e-16
v47	-0.625939	0.003930	-159.281	<2e-16
K = 8 II N = 17376 II R^2 = 0.6285				

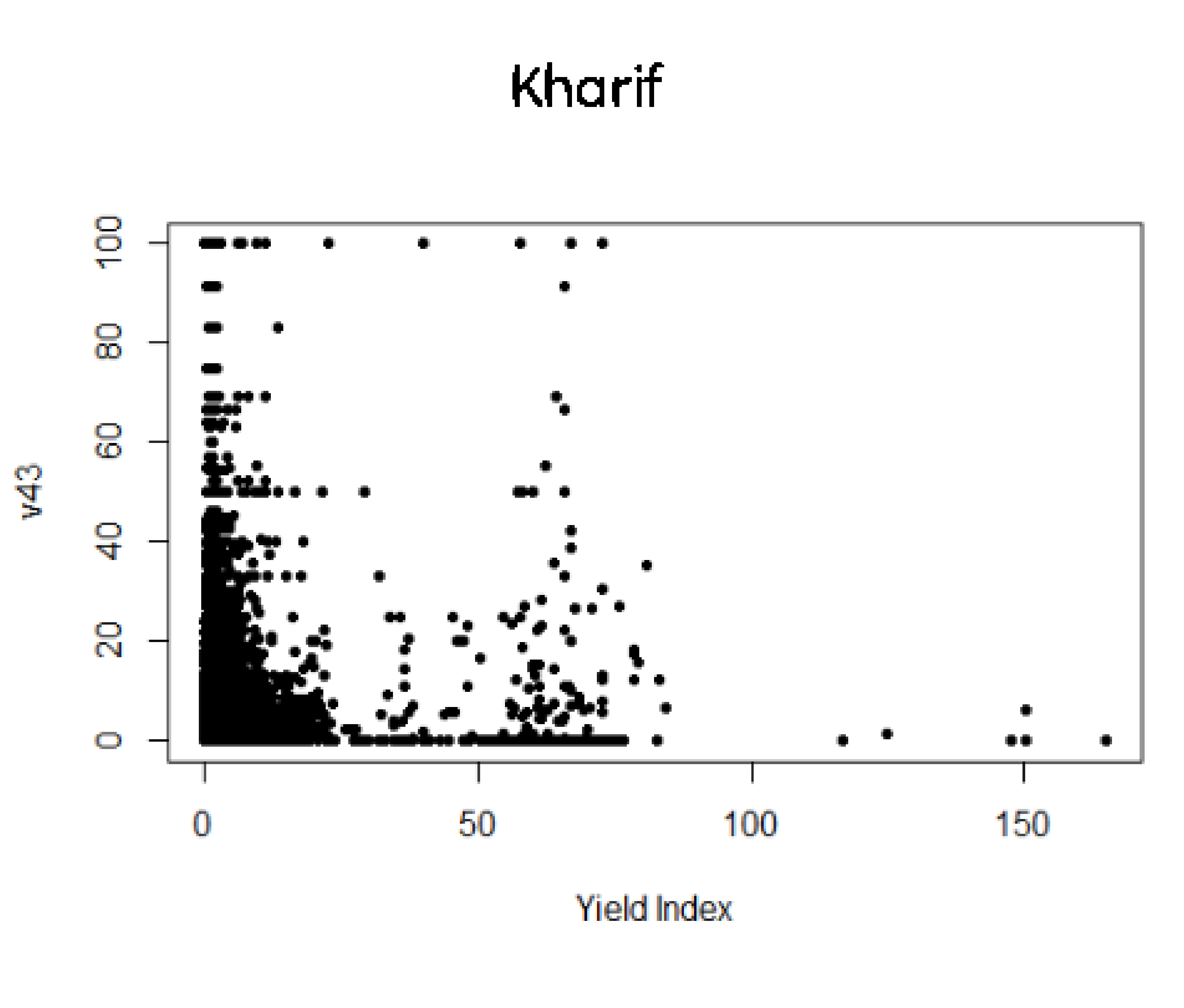
Regression Summary of Rabi

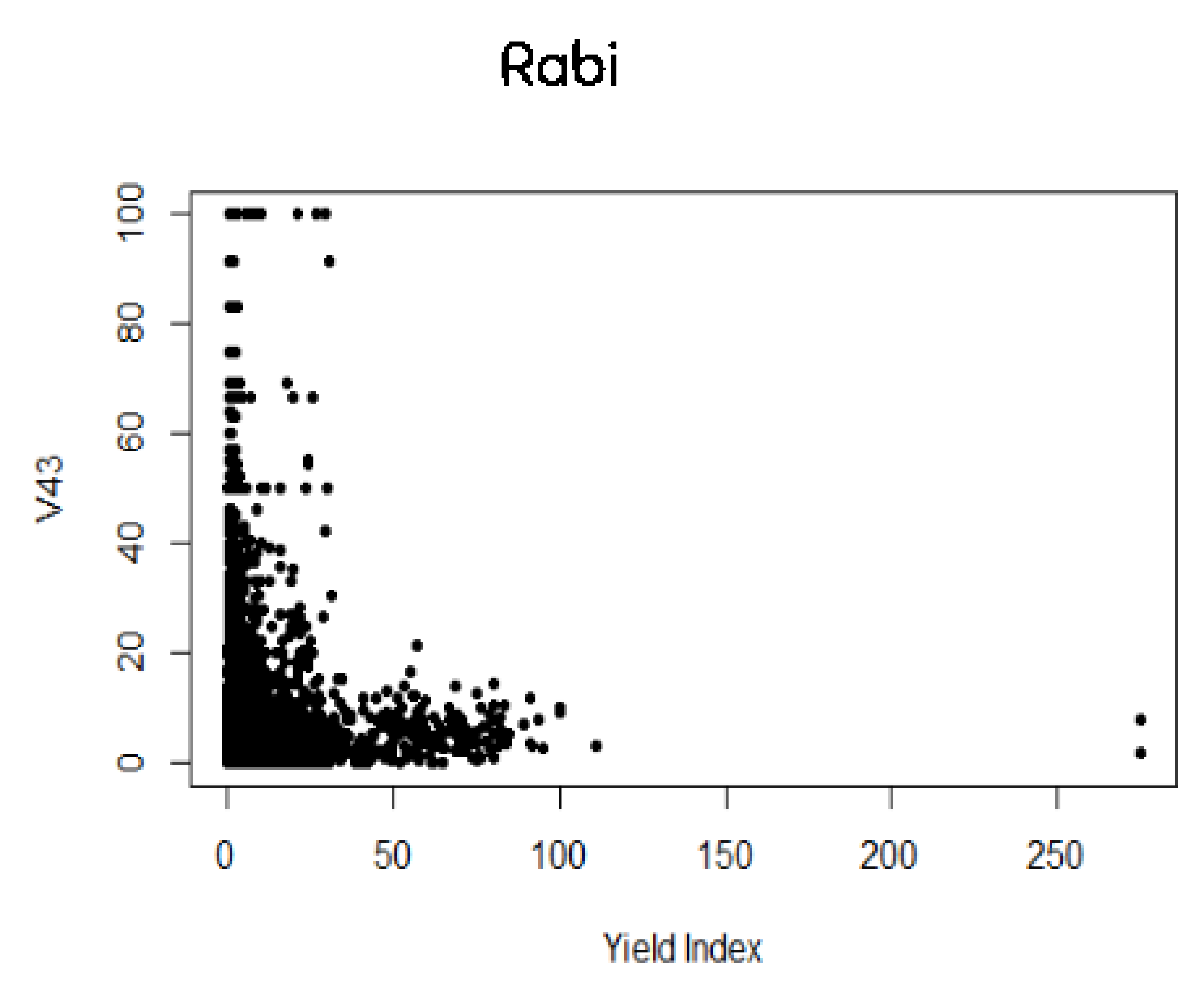
Model 2:

$$v43_{(i,t)} = \beta_0 + \beta_1 tap_{(i,t)} + \beta_2 v41_{(i,t)} + \beta_3 v42_{(i,t)} + \beta_4 v44_{(i,t)} + \beta_5 v45_{(i,t)} + \beta_6 v46_{(i,t)} + \beta_7 v47_{(i,t)}$$

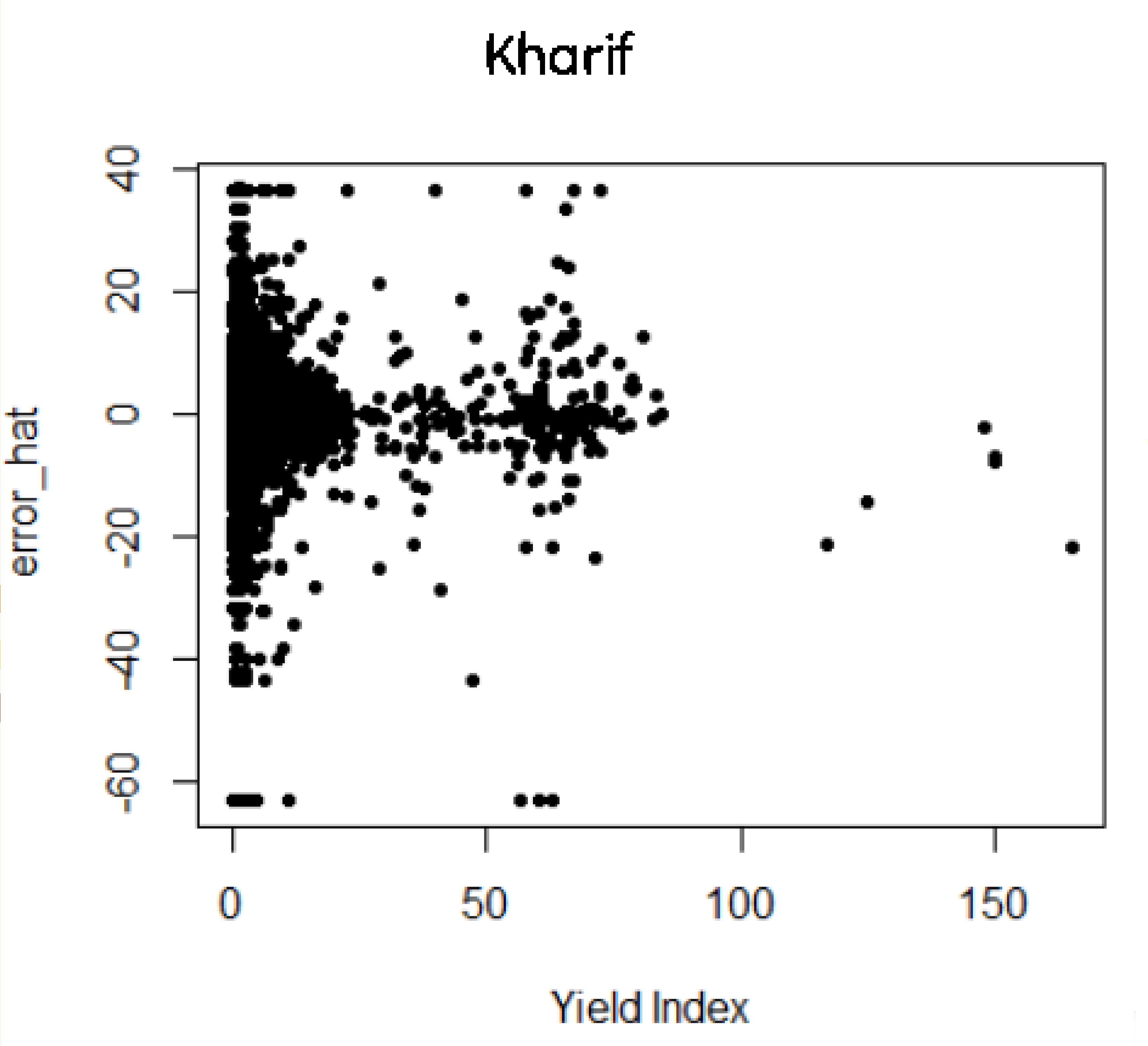
Model 1	Estimate	Standard Error	T values	P values
Intercept	63.396676	0.360804	175.709	<2e-16
tap	-0.004109	0.002219	-1.852	0.064
v41	-0.761700	0.006309	-120.724	<2e-16
v42	-0.643672	0.005214	-123.447	<2e-16
v44	-0.587252	0.009183	-63.951	<2e-16
v45	-0.581304	0.006870	-84.621	<2e-16
v46	-0.629736	0.017523	-35.938	<2e-16
v47	-0.625553	0.004153	-150.617	<2e-16
$K = 8$ || $N = 15545$ || $R^2 = 0.6254$				

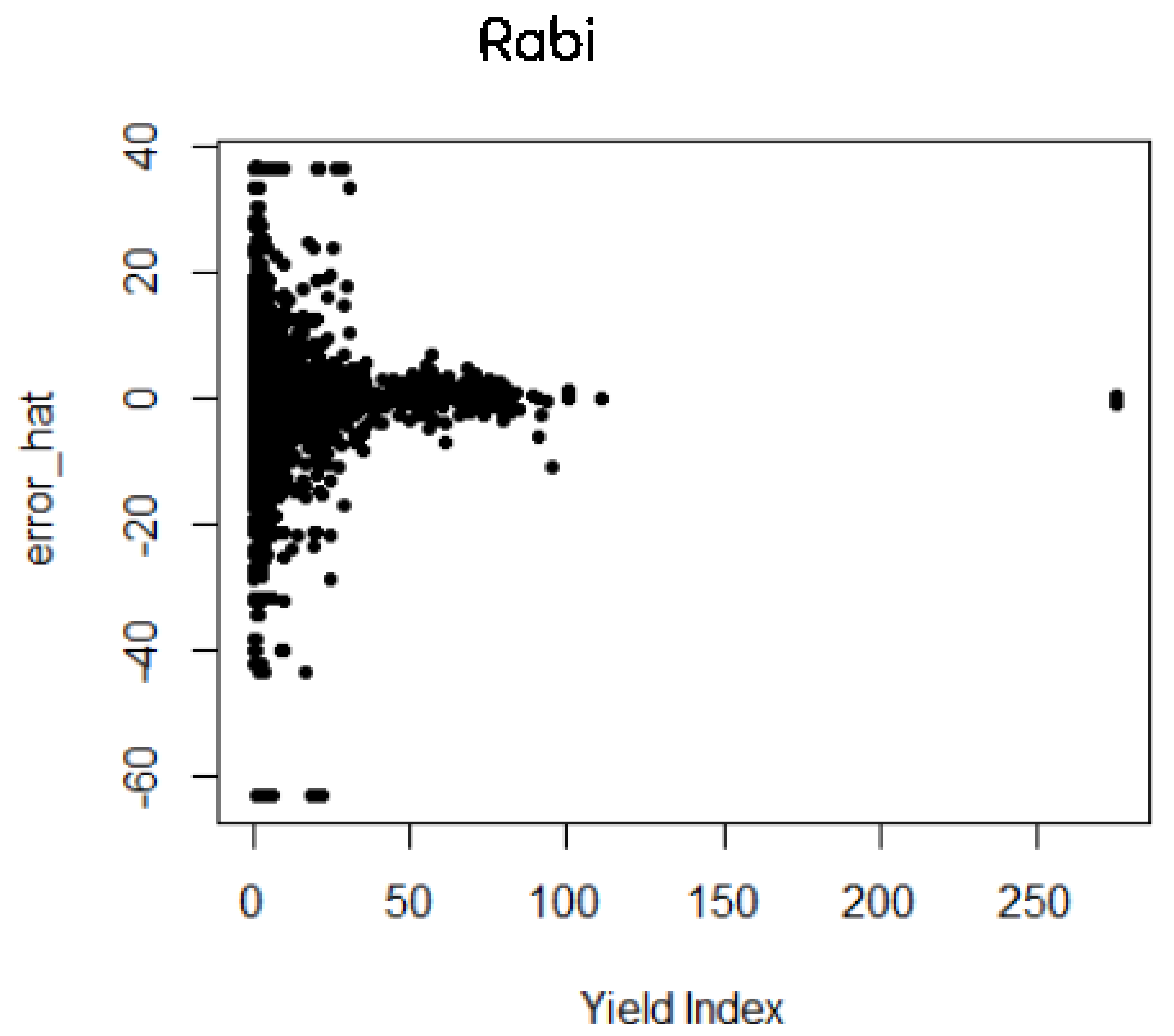
Plot of Health Indicator Vs Yield Index



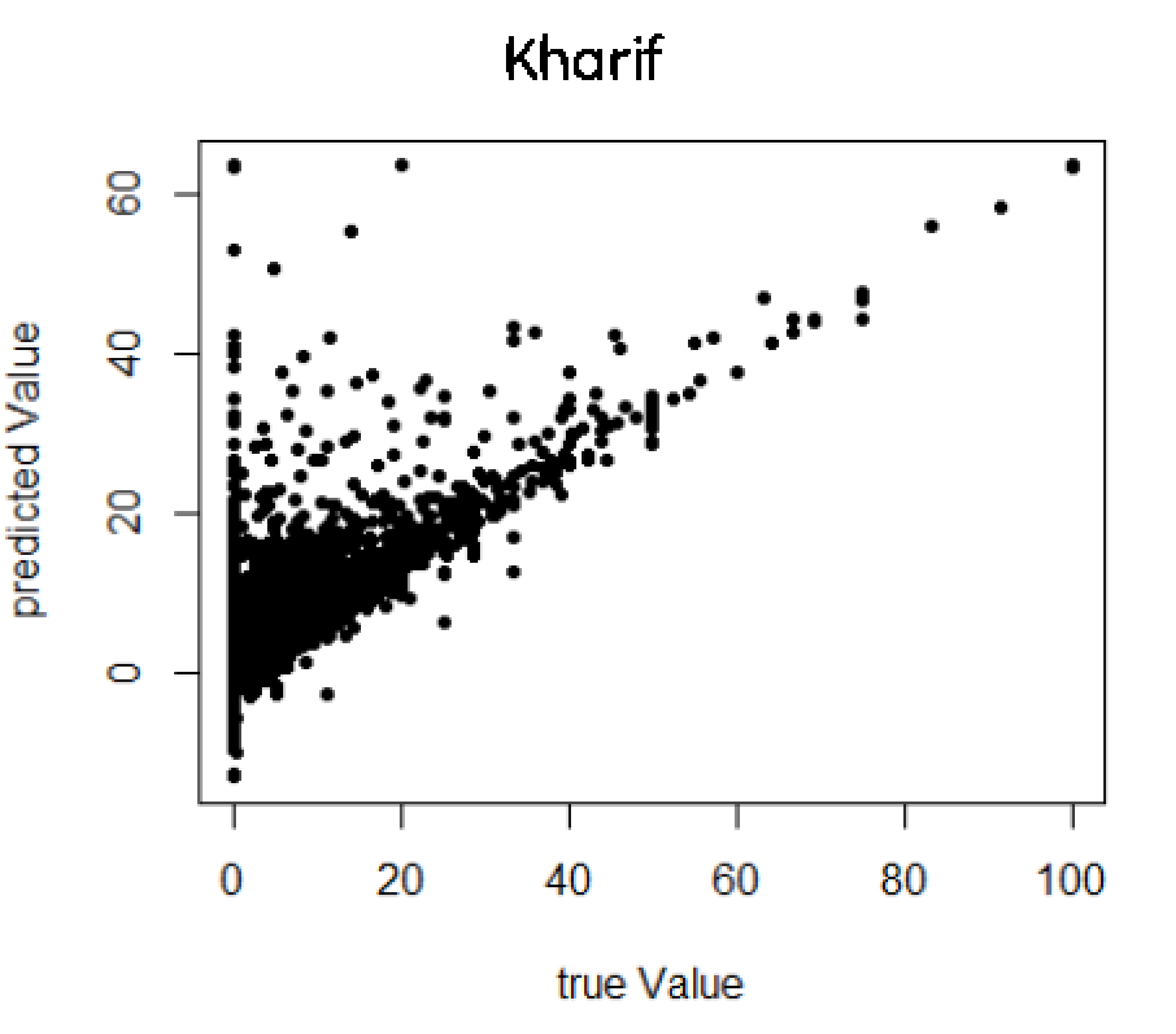


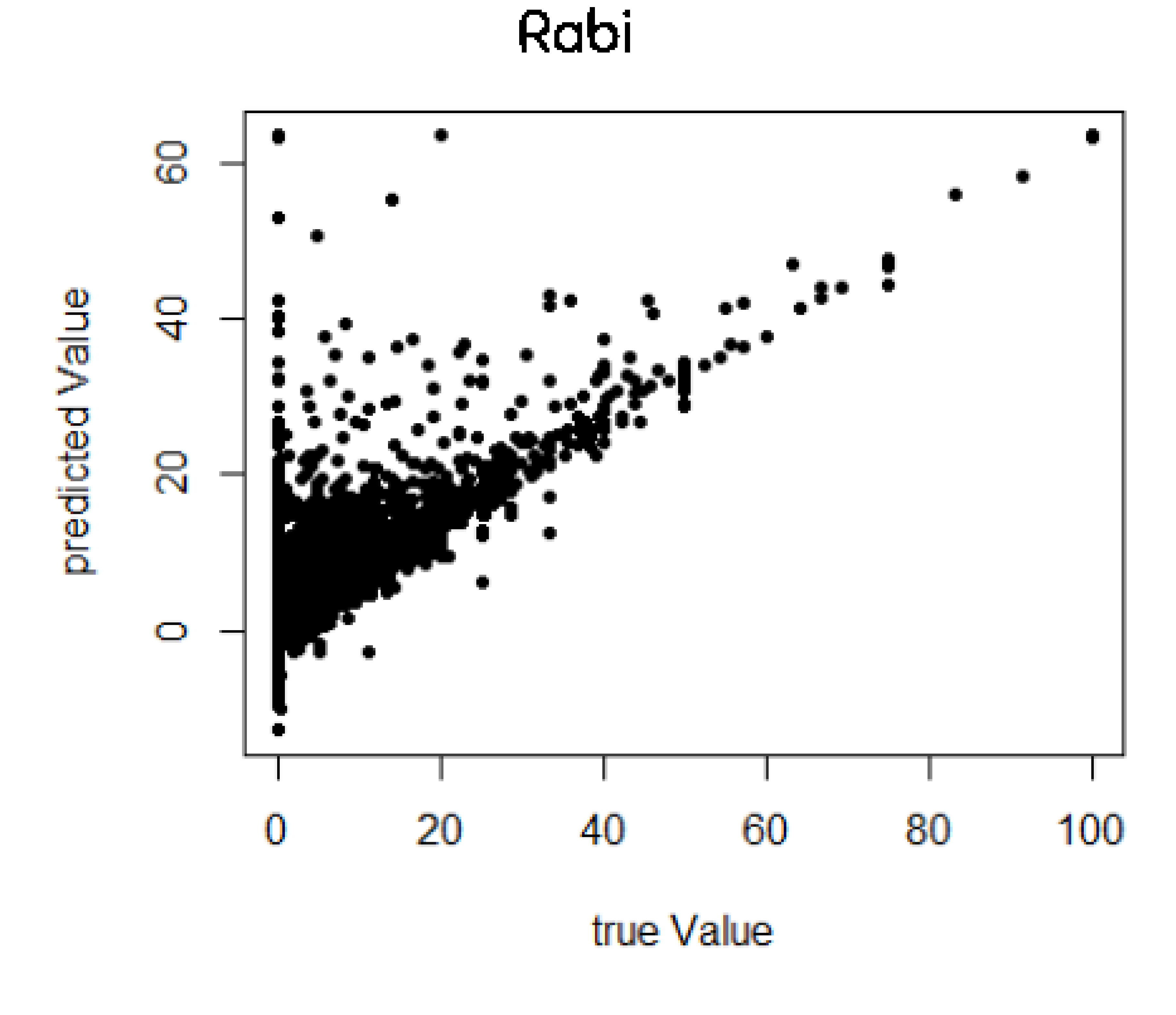
Plot of $\widehat{u_{i,t}}$ Vs Yield Index



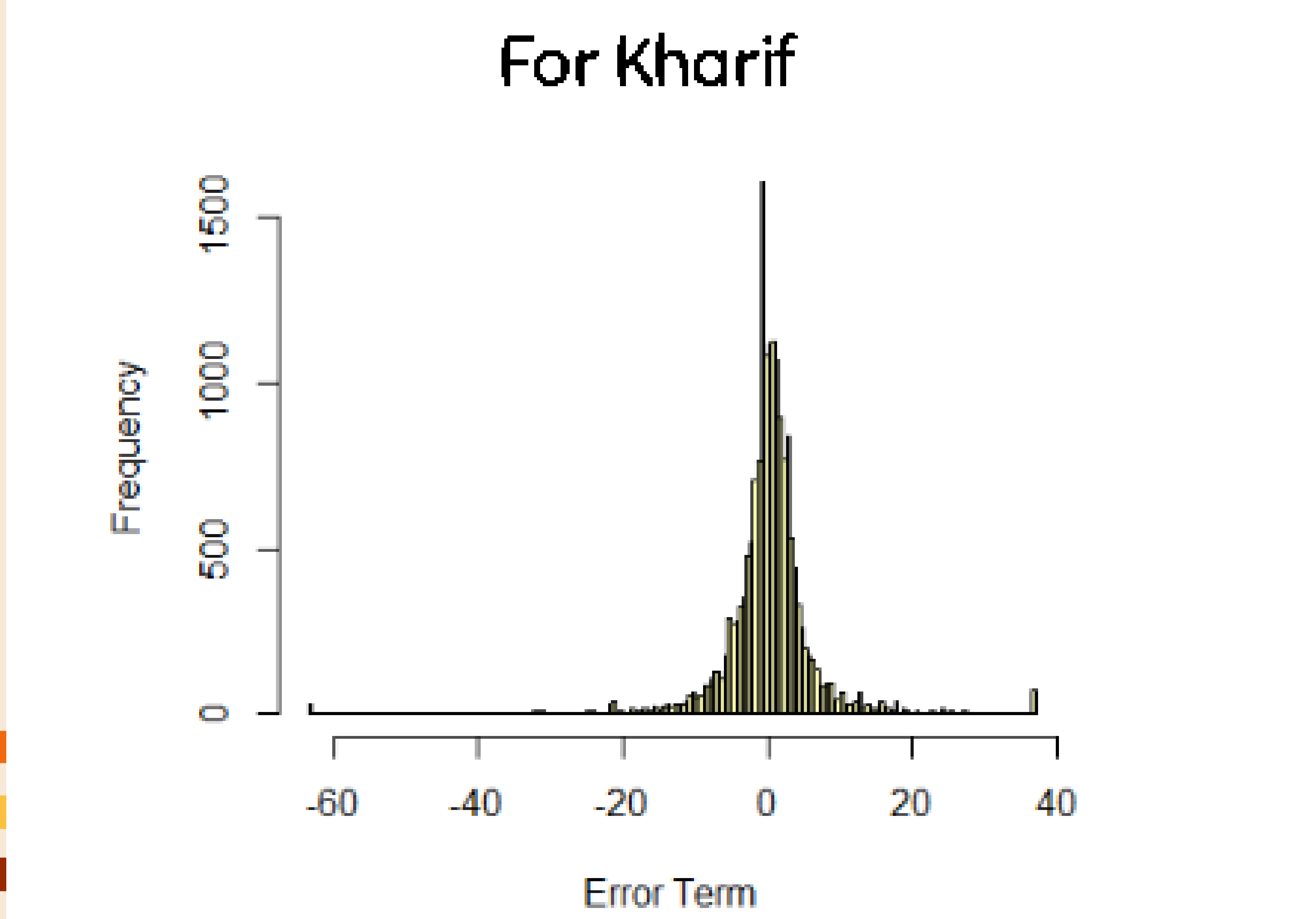


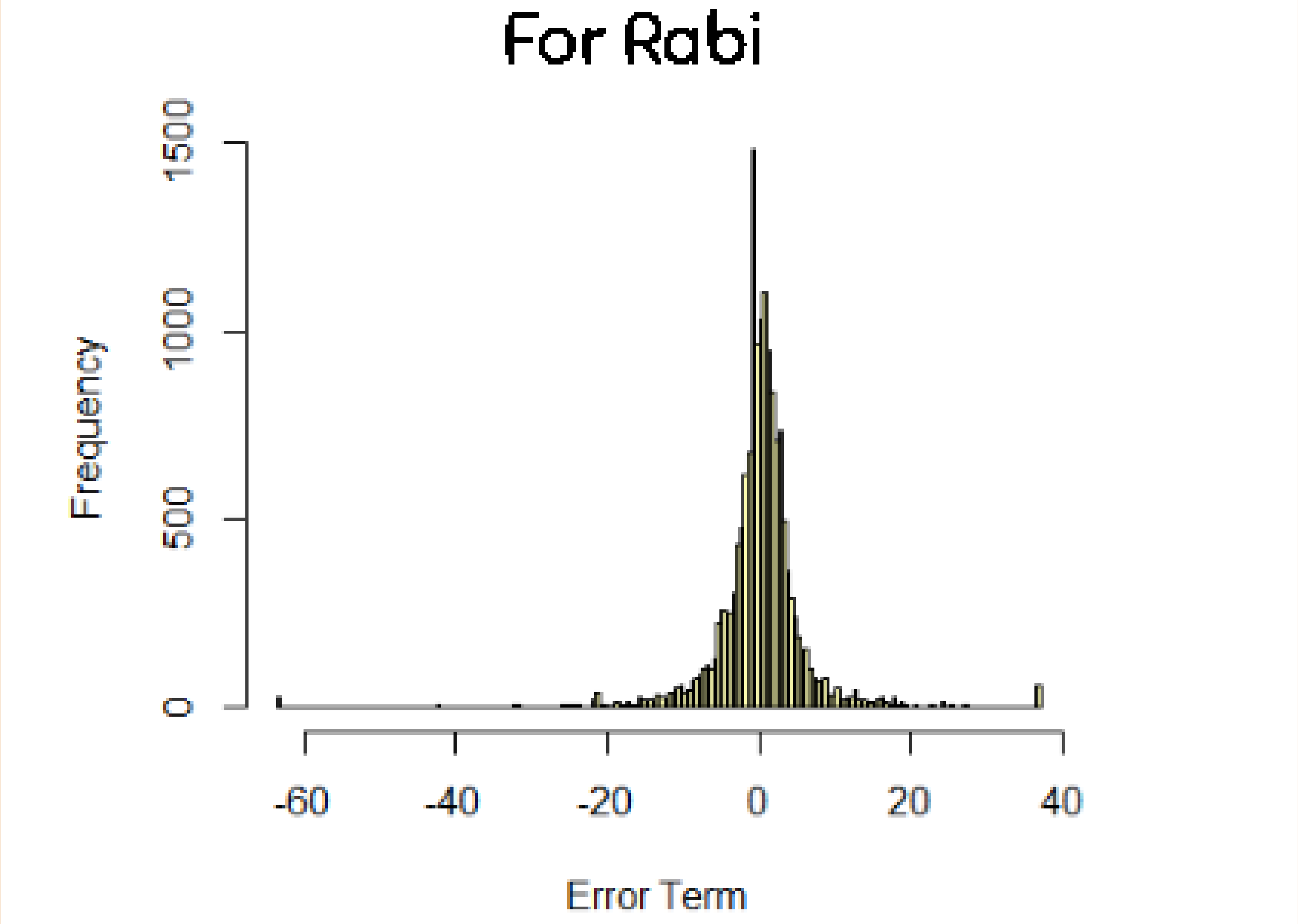
Plot of True Value Vs Expected Value





Histograms of $\widehat{u_{i,t}}$

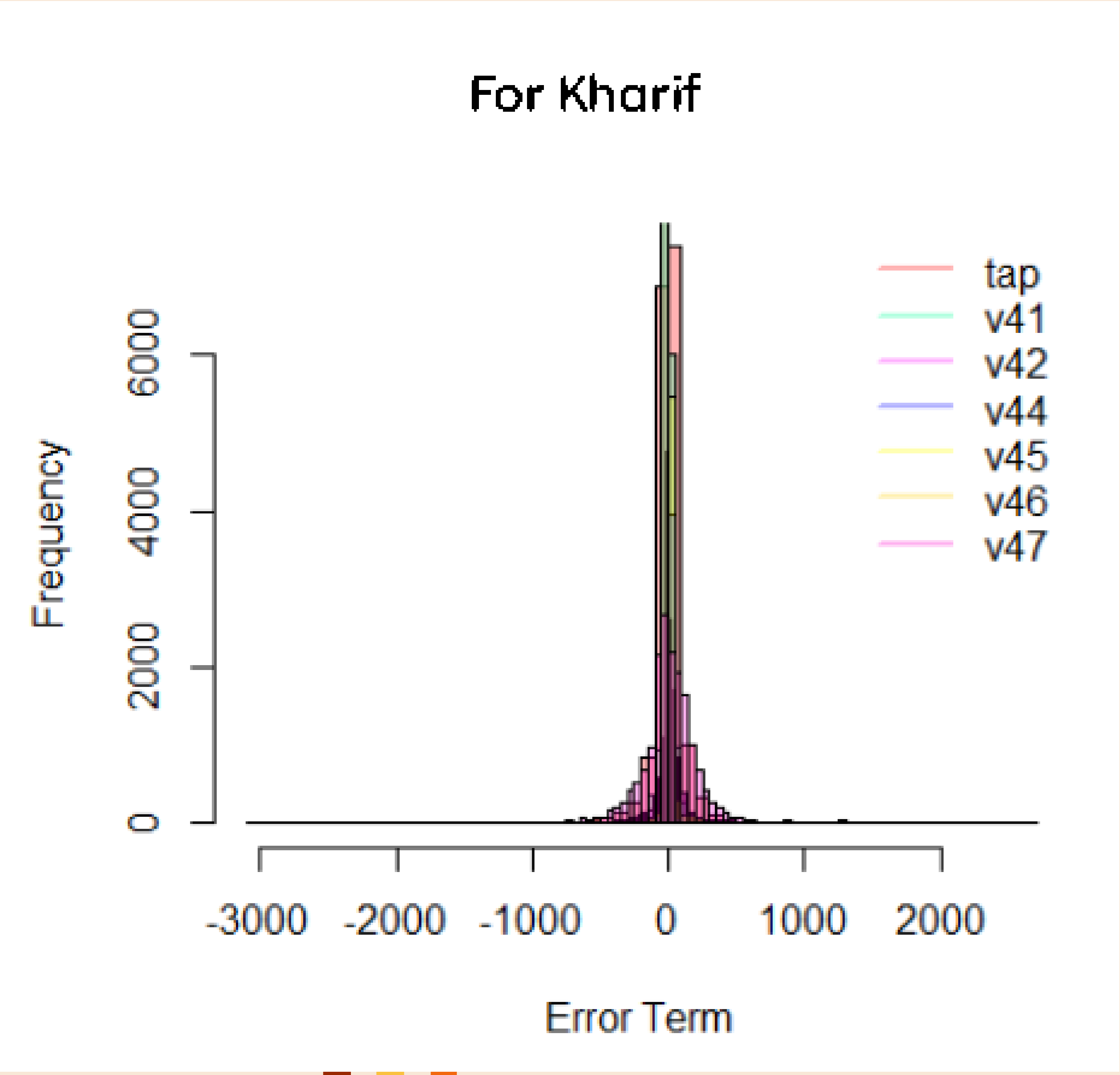


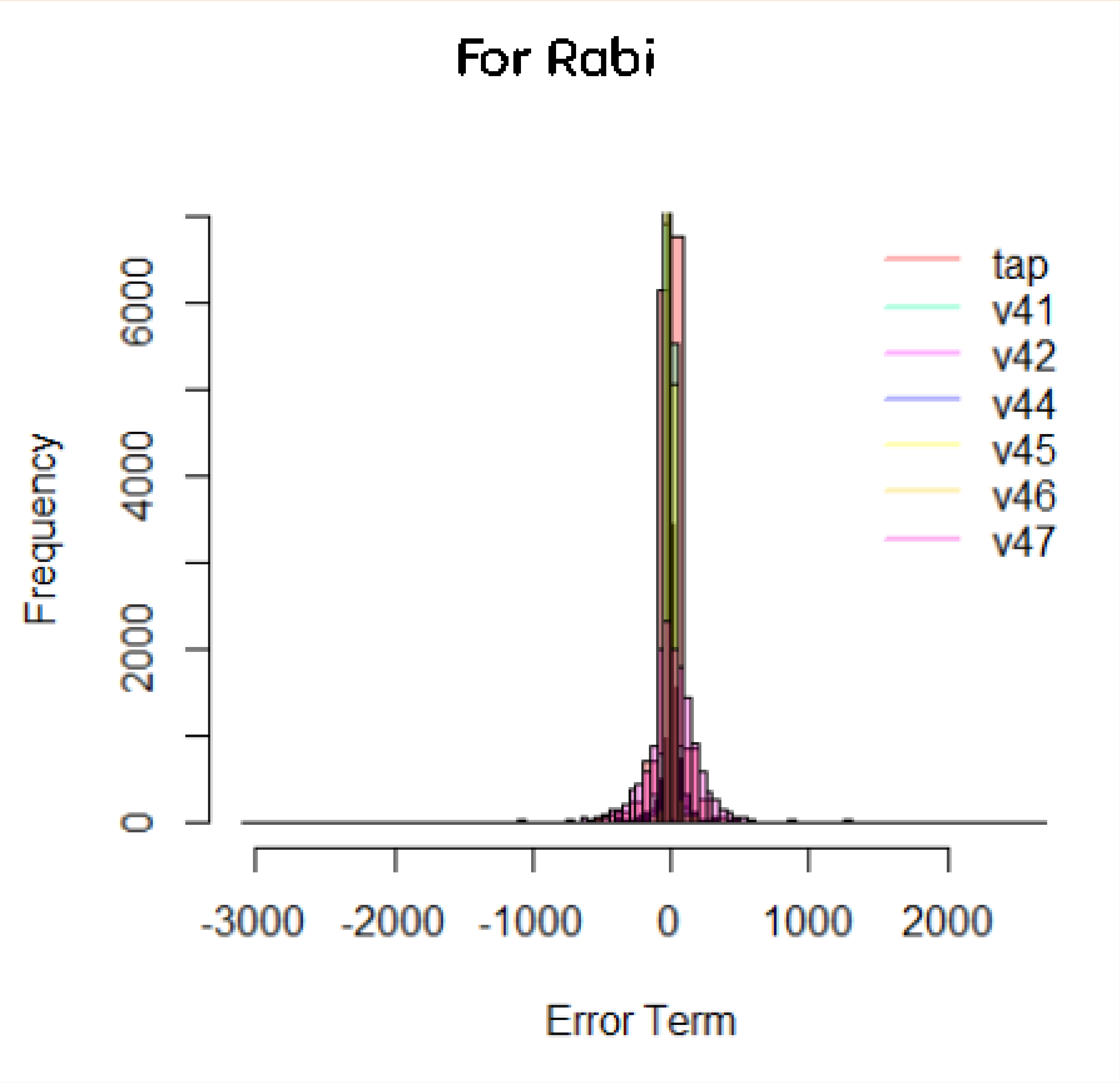



```
> error_hat_sum_kharif <- sum(error_hat_kharif,na.rm = TRUE)
> error_hat_sum_Rabi <- sum(error_hat_Rabi,na.rm = TRUE)
> error_hat_sum_kharif
[1] 2.82965e-08
> error_hat_sum_Rabi
[1] 8.735333e-09
```

The values of the summation ($\widehat{u_{i,t}}$)are all approximately equal to 0.

Histogram of $\widehat{u_{i,t}} * x_{i,t}$





$$\sum_{i,t} \widehat{u_{i,t}} x_{i,t}$$

Dependent Variable	Kharif	Rabi
tap	6.009941e-07	1.830124e-07
v41	2.973328e-07	7.741792e-08
v42	5.026667e-07	1.829574e-07
v44	5.386443e-08	1.936649e-08
v45	1.147732e-07	3.320037e-08
v46	6.045357e-09	2.544224e-09
v47	1.46376e-06	4.494355e-07

The values of the summation ($\sum_{i,t} \widehat{u_{i,t}} x_{i,t}$) are all approximately equal to 0.

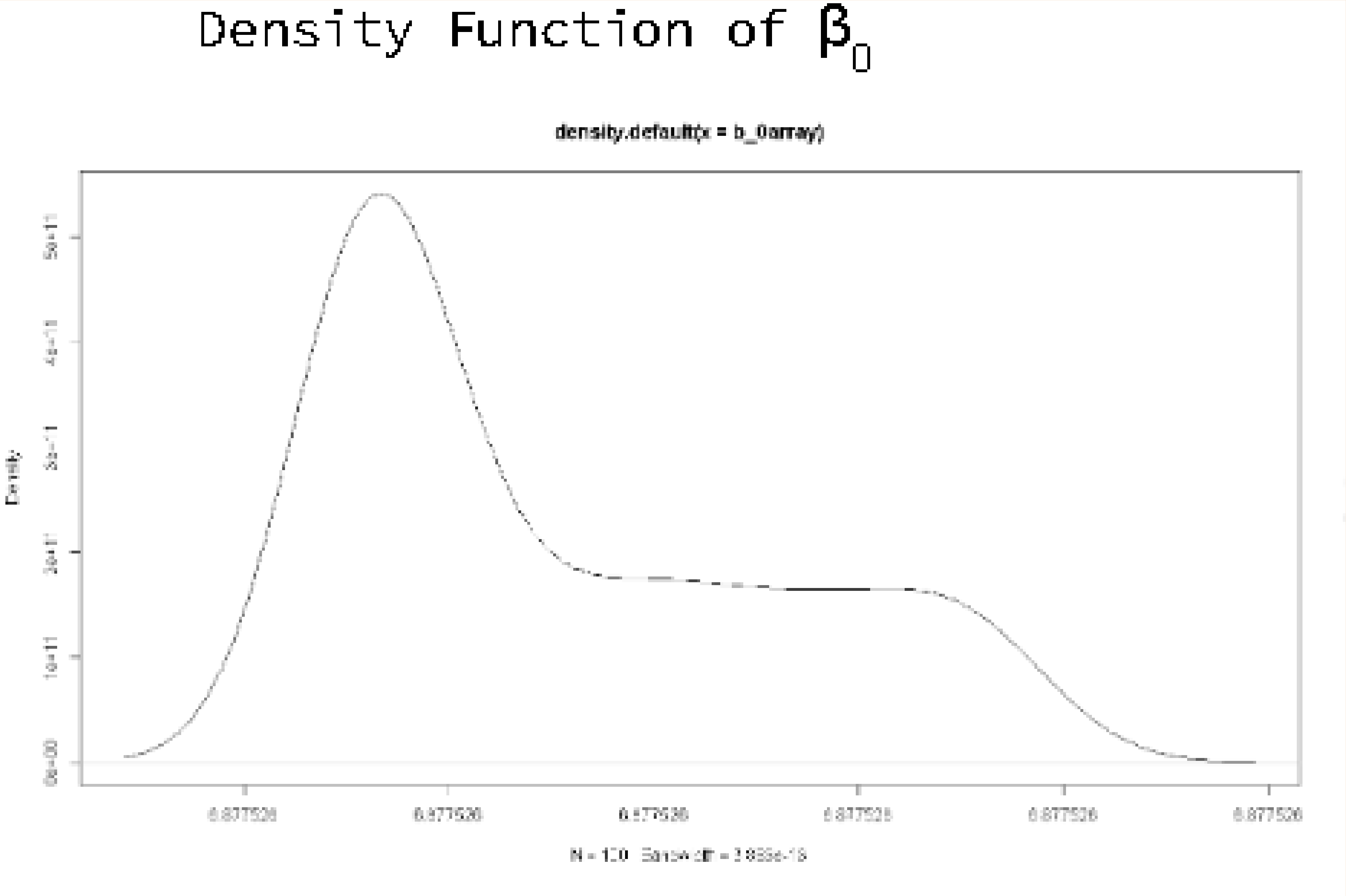
Monte Carlo Simulations:

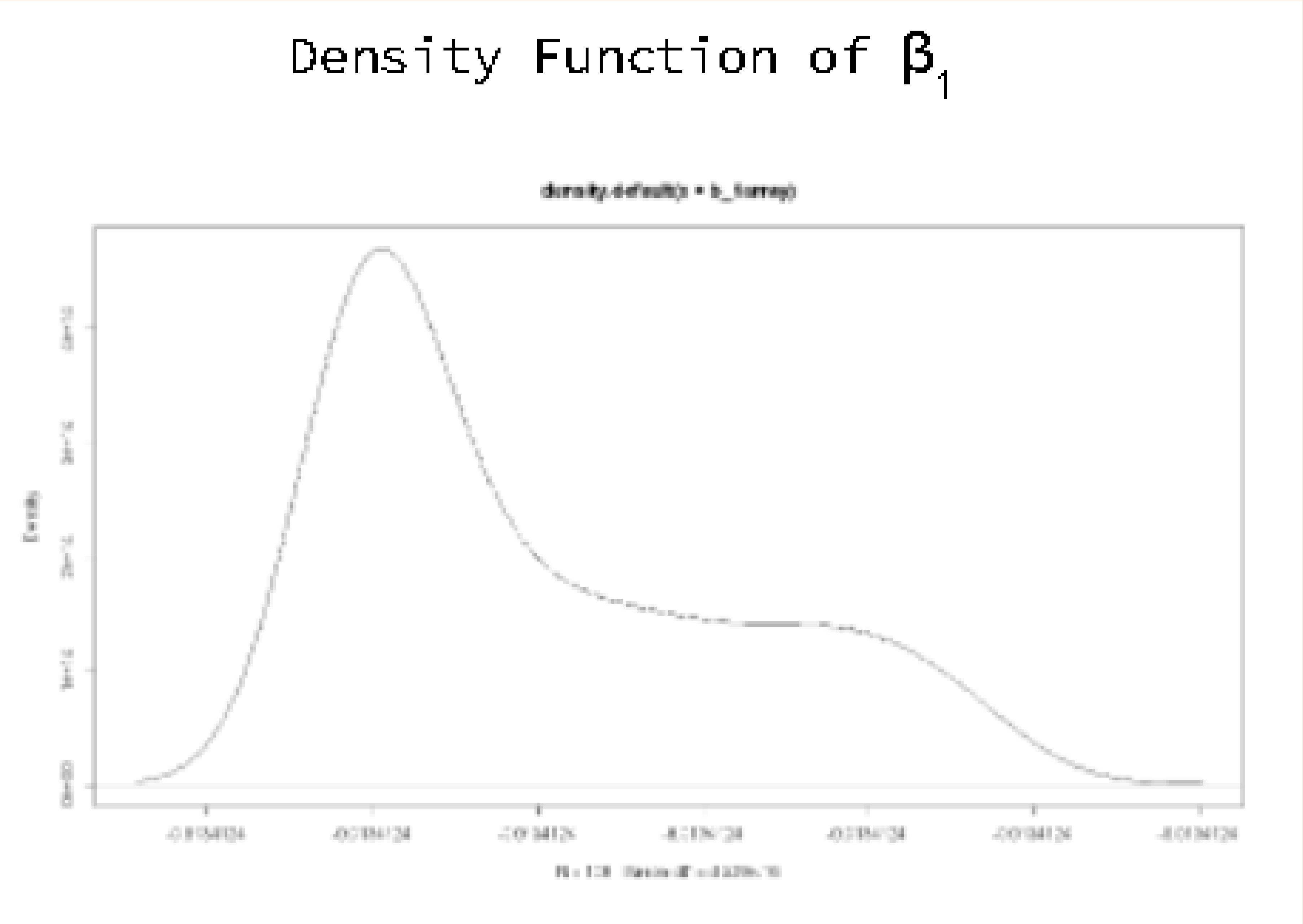
Performed Monte Carlo Simulation on v43 as dependent variable and yield index as a dependent variable after deleting 20% of the data set for the original data

Regression Model : v43(i,t) = β0 + β1index(i,t)

On complete Data Set		
	β_0	β_1
Mean	6.914854	-0.013597
Standard Error	0.053777	0.003268

On updated Data Set		
	β_0	β_1
Mean	6.877526	-0.0134124
Standard Deviation	1.083854e-12	1.208235e-15
Standard Error	1.083854e-13	1.208235e-16



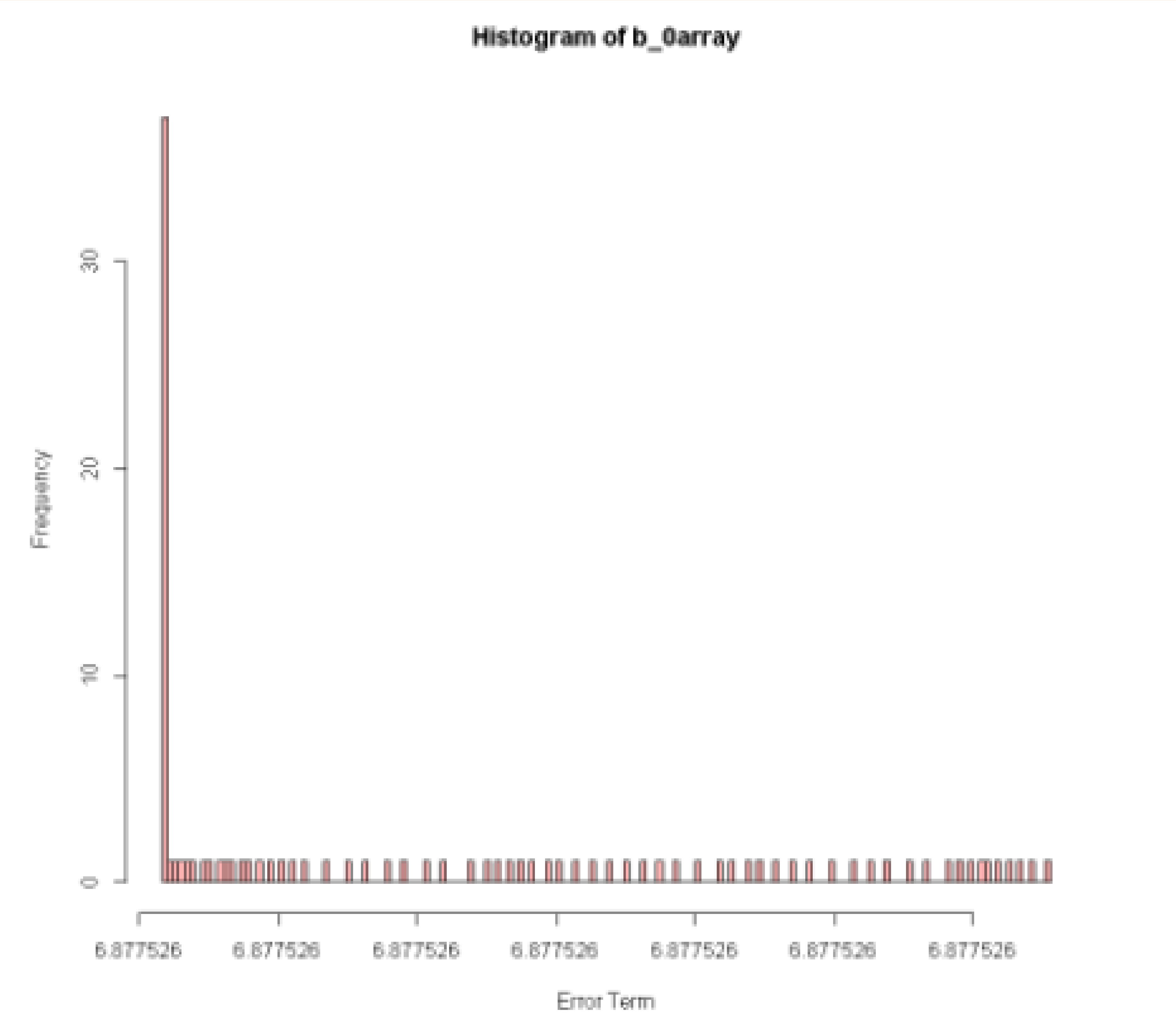


Conclusion :

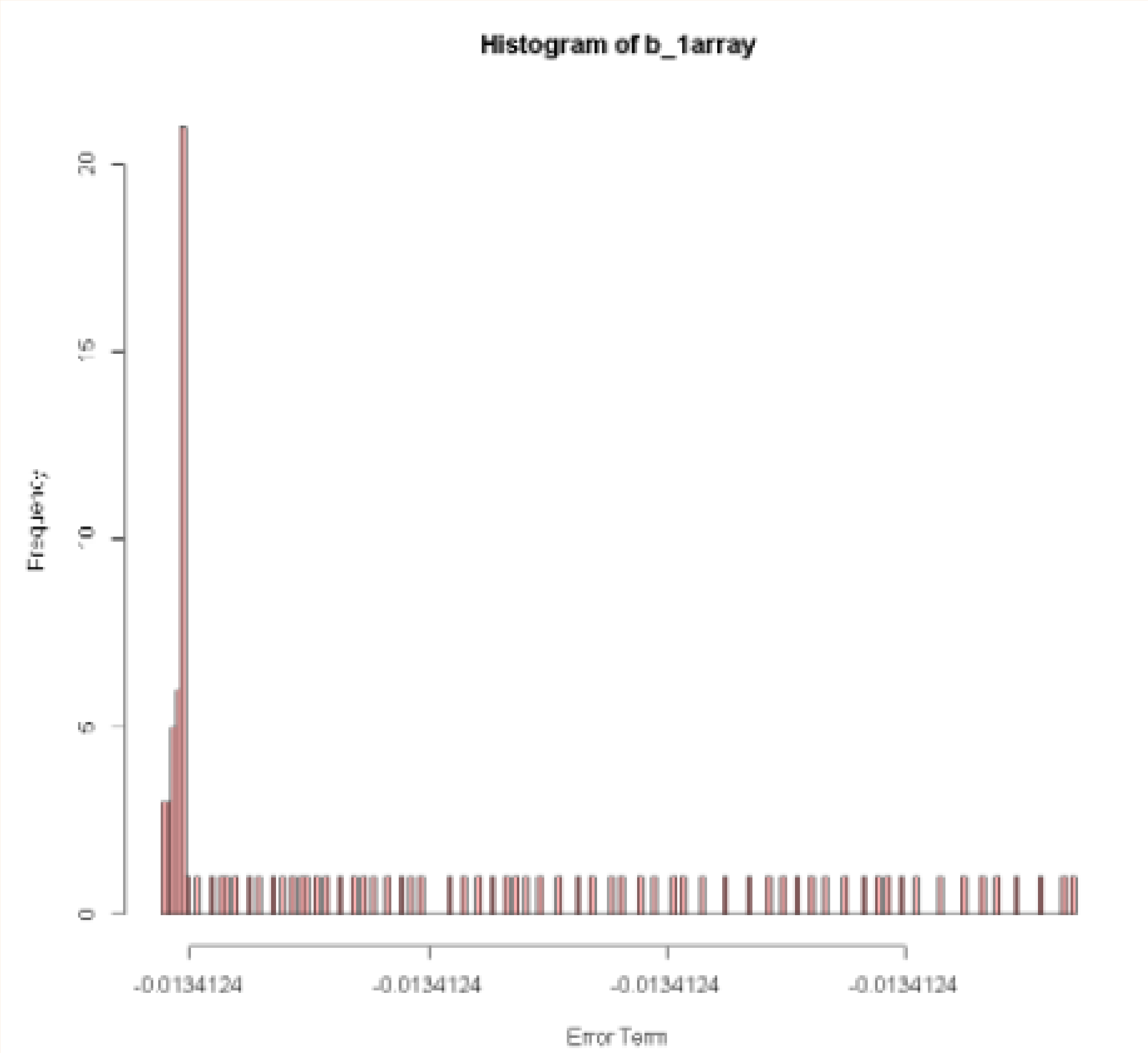
The value of β0's is within the error of 0.5398%

The value of β1's is within the error of 1.37633%

Histograms of β0



Histograms of β1



Result from Monte Carlo Simulations

- **From the table, we can deduce that the mean values of b0 and b1 are similar with some marginal errors.**
- (The error could have been improved by running the simulation for some more iterations. but to to limited machine power we were unable to reduce the error any further)
- **From the density functions, it can be derived that the peak value i.e, most of the b0 and b1 values are near the values of the b's we get from the original dataset**
- **From the histograms, We can see that the frequency is max at the values of the b's we got from the original dataset**

Hypothesis Test

i) $v43_{(i,t)} = \beta_0 + \beta_1 tap_{(i,t)} + \beta_2 v41_{(i,t)} + \beta_3 v42_{(i,t)} + \beta_4 v44_{(i,t)} + \beta_5 v45_{(i,t)} + \beta_6 v46_{(i,t)} + \beta_7 v47_{(i,t)} + \beta_8 south_{(i,t)}$

Here, south is a dummy variable which is 1 if the state is in the South Zone else 0

Kharif

```
Call:
lm(formula = v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 +
    south, data = kharif_dummy)

Residuals:
    Min      1Q  Median      3Q     Max
-63.392  -1.901   0.188   2.384  36.688

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) 63.392776   0.338215  187.434  < 2e-16 ***
tap         -0.001115   0.002175   -0.513    0.608
v41         -0.751243   0.006077 -123.617  < 2e-16 ***
v42         -0.645243   0.004981 -129.546  < 2e-16 ***
v44         -0.580344   0.008723  -66.534  < 2e-16 ***
v45         -0.582210   0.006592  -88.317  < 2e-16 ***
v46         -0.628065   0.016367  -38.374  < 2e-16 ***
v47         -0.624543   0.003929 -158.972  < 2e-16 ***
south       -0.984678   0.135553   -7.264 3.91e-13 ***
```

Rabi

```
Call:
lm(formula = v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 +
    south, data = Rabi_dummy)

Residuals:
    Min      1Q  Median      3Q     Max
-63.393  -1.789   0.209   2.348  36.611

Coefficients:
              Estimate Std. Error  t value Pr(>|t|)
(Intercept) 63.3882664  0.3599855  176.086   <2e-16 ***
tap          0.0005755  0.0022814    0.252    0.801
v41         -0.7512140  0.0064155 -117.094   <2e-16 ***
v42         -0.6443824  0.0052030 -123.849   <2e-16 ***
v44         -0.5893674  0.0091654  -64.303   <2e-16 ***
v45         -0.5834856  0.0068587  -85.072   <2e-16 ***
v46         -0.6304374  0.0174830  -36.060   <2e-16 ***
v47         -0.6242068  0.0041469 -150.524   <2e-16 ***
south       -1.1953093  0.1410030   -8.477   <2e-16 ***
```

T-test - Null Hypothesis Rejected for Kharif

T-test - Null Hypothesis Rejected for Rabi

ii) $v43_{(i,t)} = \beta_0 + \beta_1 tap_{(i,t)} + \beta_2 v41_{(i,t)} + \beta_3 v42_{(i,t)} + \beta_4 v44_{(i,t)} + \beta_5 v45_{(i,t)} + \beta_6 v46_{(i,t)} + \beta_7 v47_{(i,t)} + \beta_8 north_{(i,t)}$

Here, north is a dummy variable which is 1 if the state is in the North Zone else 0

Kharif

```
Call:
lm(formula = v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 +
    north, data = kharif_dummy)

Residuals:
    Min      1Q  Median      3Q     Max
-63.635 -1.813   0.191   2.380  36.975

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) 63.371106   0.339228  186.810   <2e-16 ***
tap         -0.004809   0.002139   -2.249   0.0246 *
v41         -0.759667   0.005966 -127.339   <2e-16 ***
v42         -0.643308   0.005021 -128.111   <2e-16 ***
v44         -0.579147   0.008735  -66.304   <2e-16 ***
v45         -0.581798   0.006630  -87.748   <2e-16 ***
v46         -0.629307   0.016414  -38.339   <2e-16 ***
v47         -0.626379   0.003936 -159.127   <2e-16 ***
north        0.270416   0.142740    1.894   0.0582 .
```

Rabi

```
Call:
lm(formula = v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 +
    north, data = Rabi_dummy)

Residuals:
    Min      1Q  Median      3Q     Max
-63.731 -1.713   0.203   2.331  37.003

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) 63.335275   0.361421  175.240  < 2e-16 ***
tap         -0.004700   0.002228   -2.109  0.03494 *
v41         -0.760837   0.006316 -120.463  < 2e-16 ***
v42         -0.641894   0.005253 -122.193  < 2e-16 ***
v44         -0.587938   0.009184  -64.015  < 2e-16 ***
v45         -0.583202   0.006903  -84.488  < 2e-16 ***
v46         -0.632086   0.017540  -36.037  < 2e-16 ***
v47         -0.626124   0.004158 -150.597  < 2e-16 ***
north        0.402155   0.146511    2.745  0.00606 **
```

T-test - failed to reject null Hypothesis for Kharif

T-test - Null Hypothesis Rejected for Rabi

iii) $v43_{(i,t)} = v43_{(i,t)} = \beta_0 + \beta_1 tap_{(i,t)} + \beta_2 v41_{(i,t)} + \beta_3 v42_{(i,t)} + \beta_4 v44_{(i,t)} + \beta_5 v45_{(i,t)} + \beta_6 v46_{(i,t)} + \beta_7 v47_{(i,t)} + \beta_8 east_{(i,t)}$

Here, east is a dummy variable which is 1 if the state is in the East Zone else 0

Kharif

```
Call:
lm(formula = v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 +
    east, data = kharif_dummy)

Residuals:
    Min      1Q  Median      3Q     Max
-63.429  -1.802   0.190   2.365  37.176

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) 63.438609   0.338700  187.301  < 2e-16 ***
tap         -0.006939   0.002236   -3.103 0.001922 **
v41         -0.758499   0.005975 -126.939  < 2e-16 ***
v42         -0.642687   0.005008 -128.336  < 2e-16 ***
v44         -0.577316   0.008738  -66.067  < 2e-16 ***
v45         -0.580194   0.006597  -87.954  < 2e-16 ***
v46         -0.626514   0.016388  -38.230  < 2e-16 ***
v47         -0.625091   0.003935 -158.845  < 2e-16 ***
east        -0.591628   0.161802   -3.656 0.000256 ***
```

Rabi

```
Call:
lm(formula = v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 +
    east, data = Rabi_dummy)

Residuals:
    Min      1Q  Median      3Q     Max
-63.411  -1.712   0.190   2.316  36.987

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) 63.419197   0.361013  175.670   <2e-16 ***
tap         -0.005637   0.002386   -2.362   0.0182 *
v41         -0.760888   0.006326 -120.274   <2e-16 ***
v42         -0.642675   0.005245 -122.527   <2e-16 ***
v44         -0.586639   0.009189  -63.841   <2e-16 ***
v45         -0.581146   0.006870  -84.596   <2e-16 ***
v46         -0.629320   0.017523  -35.914   <2e-16 ***
v47         -0.624990   0.004166 -150.036   <2e-16 ***
east        -0.252251   0.145079   -1.739   0.0821 .
```

T-test - Null Hypothesis Rejected for Kharif

T-test - failed to reject null Hypothesis for Rabi

iv) $v43_{(i,t)} = \beta_0 + \beta_1 tap_{(i,t)} + \beta_2 v41_{(i,t)} + \beta_3 v42_{(i,t)} + \beta_4 v44_{(i,t)} + \beta_5 v45_{(i,t)} + \beta_6 v46_{(i,t)} + \beta_7 v47_{(i,t)} + \beta_8 west_{(i,t)}$

Here, west is a dummy variable which is 1 if the state is in the West Zone else zero

Kharif

```
Call:
lm(formula = v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 +
    west, data = kharif_dummy)

Residuals:
    Min      1Q  Median      3Q     Max
-63.423  -1.852   0.182   2.360  36.614

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) 63.385109   0.337759  187.664   <2e-16 ***
tap          0.001652   0.002209    0.748    0.455
v41         -0.761069   0.005942 -128.084   <2e-16 ***
v42         -0.636048   0.005043 -126.126   <2e-16 ***
v44         -0.578700   0.008708  -66.456   <2e-16 ***
v45         -0.583048   0.006584  -88.553   <2e-16 ***
v46         -0.628257   0.016345  -38.438   <2e-16 ***
v47         -0.625370   0.003919 -159.572   <2e-16 ***
west        -1.410486   0.141069   -9.999   <2e-16 ***
```

Rabi

```
Call:
lm(formula = v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 +
    west, data = Rabi_dummy)

Residuals:
    Min      1Q  Median      3Q     Max
-63.401  -1.751   0.221   2.304  36.631

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) 63.368523   0.360131  175.960  < 2e-16 ***
tap          0.001407   0.002325    0.605    0.545
v41         -0.762450   0.006298 -121.060  < 2e-16 ***
v42         -0.637556   0.005263 -121.138  < 2e-16 ***
v44         -0.587244   0.009165  -64.072  < 2e-16 ***
v45         -0.583022   0.006860  -84.990  < 2e-16 ***
v46         -0.630047   0.017489  -36.025  < 2e-16 ***
v47         -0.625261   0.004145 -150.829  < 2e-16 ***
west        -1.229806   0.157869   -7.790 7.12e-15 ***
```

T-test - Null Hypothesis Rejected for Kharif

T-test - Null Hypothesis Rejected for Rabi

v) $v43_{(i,t)} = \beta_0 + \beta_1 tap_{(i,t)} + \beta_2 v41_{(i,t)} + \beta_3 v42_{(i,t)} + \beta_4 v44_{(i,t)} + \beta_5 v45_{(i,t)} + \beta_6 v46_{(i,t)} + \beta_7 v47_{(i,t)} + \beta_8 central_{(i,t)}$

Here, central is a dummy variable which is 1 if the state is in the Central Zone else 0

Kharif

```
Call:
lm(formula = v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 +
    central, data = kharif_dummy)

Residuals:
    Min      1Q  Median      3Q     Max
-63.304  -1.833   0.161   2.353  36.891

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) 63.305970   0.339948  186.223  < 2e-16 ***
tap         -0.002741   0.002188   -1.253  0.21038
v41         -0.757456   0.006014 -125.951  < 2e-16 ***
v42         -0.646875   0.005038 -128.396  < 2e-16 ***
v44         -0.577271   0.008741  -66.041  < 2e-16 ***
v45         -0.579081   0.006611  -87.598  < 2e-16 ***
v46         -0.626549   0.016389  -38.229  < 2e-16 ***
v47         -0.625922   0.003929 -159.324  < 2e-16 ***
central      0.551437   0.163753    3.367  0.00076 ***
```

Rabi

```
Call:
lm(formula = v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 +
    central, data = Rabi_dummy)

Residuals:
    Min      1Q  Median      3Q     Max
-63.300  -1.762   0.196   2.292  36.900

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) 63.302272   0.362259  174.743  < 2e-16 ***
tap         -0.002807   0.002265   -1.239  0.21529
v41         -0.759192   0.006370 -119.185  < 2e-16 ***
v42         -0.645526   0.005254 -122.866  < 2e-16 ***
v44         -0.585966   0.009192  -63.747  < 2e-16 ***
v45         -0.579936   0.006885  -84.233  < 2e-16 ***
v46         -0.628769   0.017522  -35.885  < 2e-16 ***
v47         -0.625372   0.004153 -150.589  < 2e-16 ***
central      0.499444   0.176321    2.833  0.00462 **
```

T-test - Null Hypothesis Rejected for kharif

T-test - Null Hypothesis Rejected for Rabi

vi) $v43_{(i,t)} = \beta_0 + \beta_1 tap_{(i,t)} + \beta_2 v41_{(i,t)} + \beta_3 v42_{(i,t)} + \beta_4 v44_{(i,t)} + \beta_5 v45_{(i,t)} + \beta_6 v46_{(i,t)} + \beta_7 v47_{(i,t)} + \beta_8 northeast_{(i,t)}$

Here, northwest is a dummy variable which is 1 if the state is in the North East Zone else 0

Kharif

```
Call:
lm(formula = v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 +
    northEast, data = kharif_dummy)

Residuals:
    Min      1Q  Median      3Q     Max
-61.616  -1.878   0.124   2.232  38.499

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) 61.498697   0.343354  179.112   <2e-16 ***
tap          0.005061   0.002136    2.369   0.0178 *
v41         -0.749543   0.005885 -127.373   <2e-16 ***
v42         -0.620147   0.005019 -123.561   <2e-16 ***
v44         -0.582203   0.008600  -67.698   <2e-16 ***
v45         -0.589894   0.006509  -90.626   <2e-16 ***
v46         -0.613785   0.016150  -38.004   <2e-16 ***
v47         -0.613654   0.003905 -157.149   <2e-16 ***
northEast    3.597762   0.153848   23.385   <2e-16 ***
```

Rabi

```
Call:
lm(formula = v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 +
    northEast, data = kharif_dummy)

Residuals:
    Min      1Q  Median      3Q     Max
-61.616  -1.878   0.124   2.232  38.499

Coefficients:
             Estimate Std. Error  t value Pr(>|t|)
(Intercept) 61.498697   0.343354  179.112   <2e-16 ***
tap          0.005061   0.002136    2.369   0.0178 *
v41         -0.749543   0.005885 -127.373   <2e-16 ***
v42         -0.620147   0.005019 -123.561   <2e-16 ***
v44         -0.582203   0.008600  -67.698   <2e-16 ***
v45         -0.589894   0.006509  -90.626   <2e-16 ***
v46         -0.613785   0.016150  -38.004   <2e-16 ***
v47         -0.613654   0.003905 -157.149   <2e-16 ***
northEast    3.597762   0.153848   23.385   <2e-16 ***
```

T-test - Null Hypothesis Rejected for kharif
T-test - Null Hypothesis Rejected for Rabi

F -Test

Kharif :

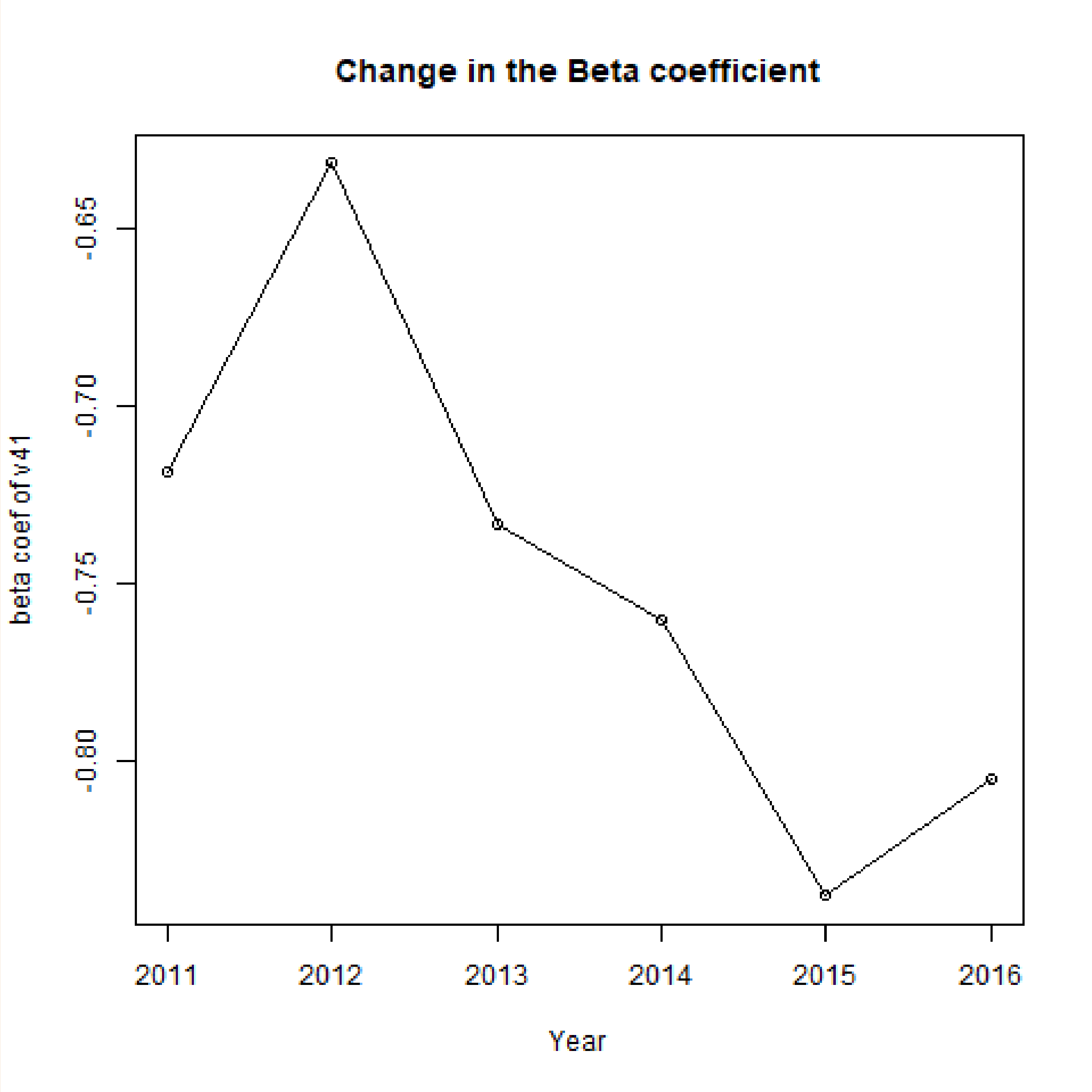
```
Model 1: v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 + south
Model 2: v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 + north + east +
    west + northEast + central
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1  17367 800190
2  17363 767823  4     32367 182.98 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Analysis of Variance Table

Model 1: v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 + south
Model 2: v43 ~ tap + v41 + v42 + v44 + v45 + v46 + v47 + north + east +
    west + northEast + central
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1  15536 686045
2  15532 658131  4     27914 164.69 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

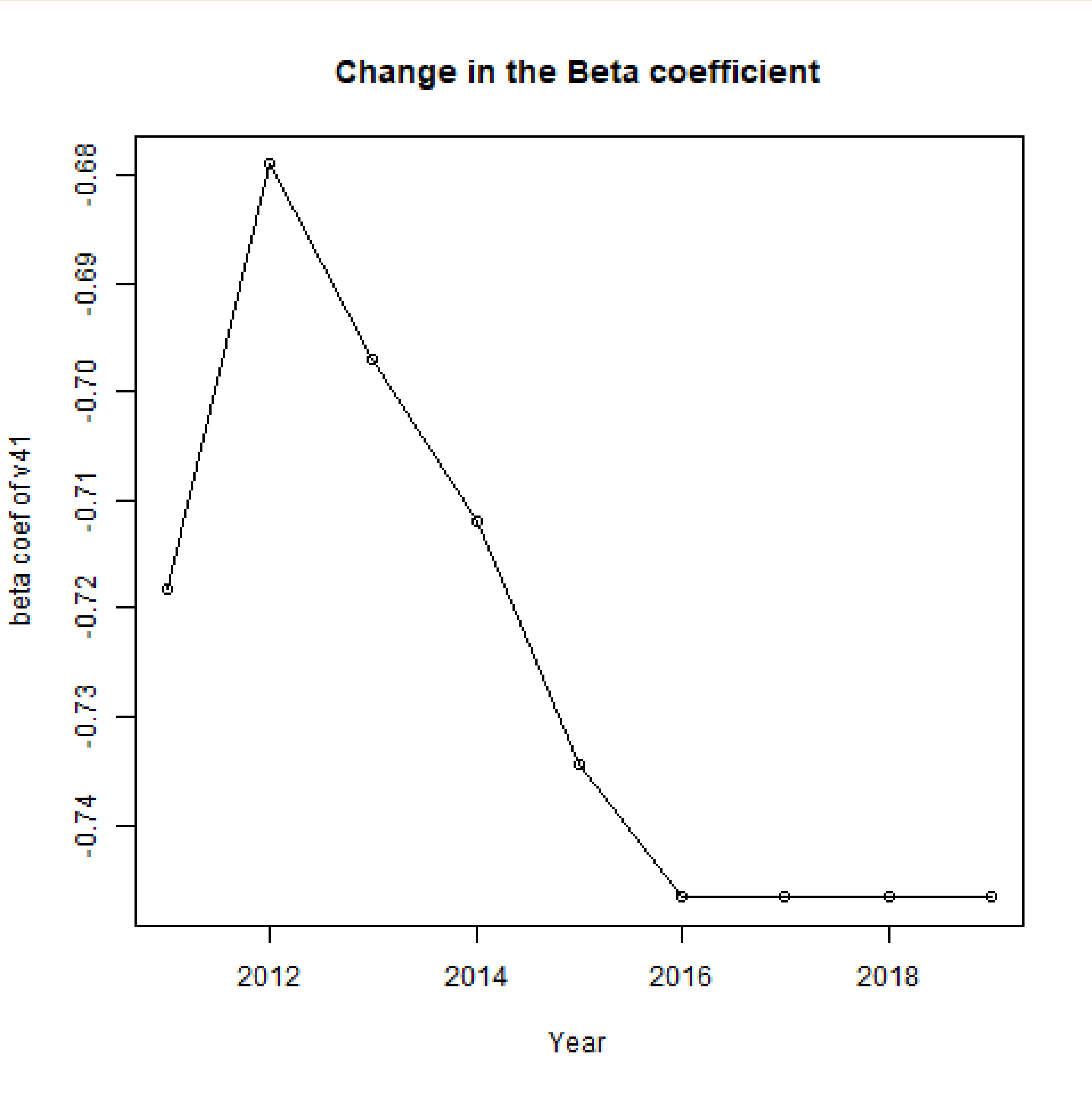
: Rabi

Creative Component



year wise

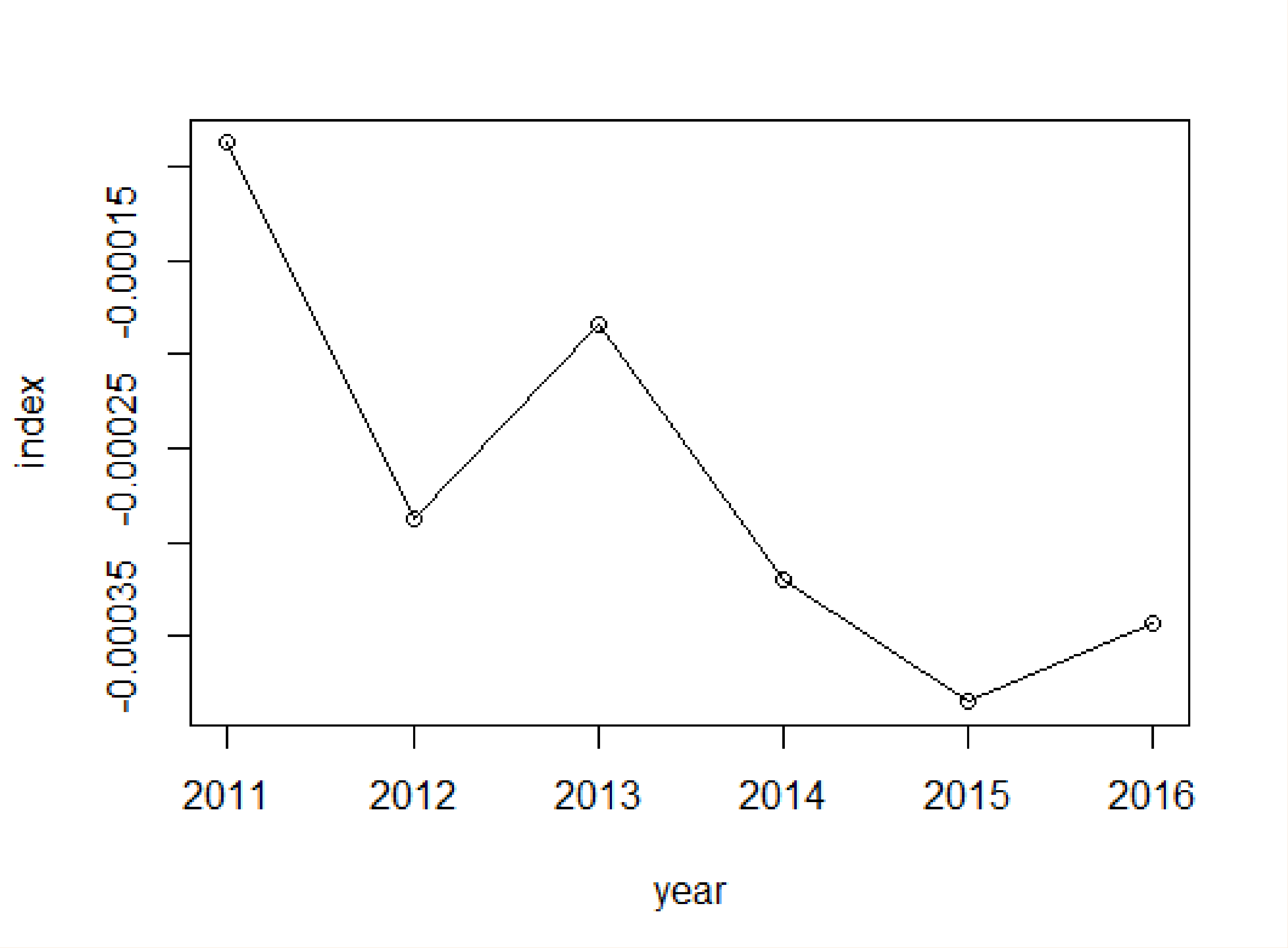
Accumulative wise



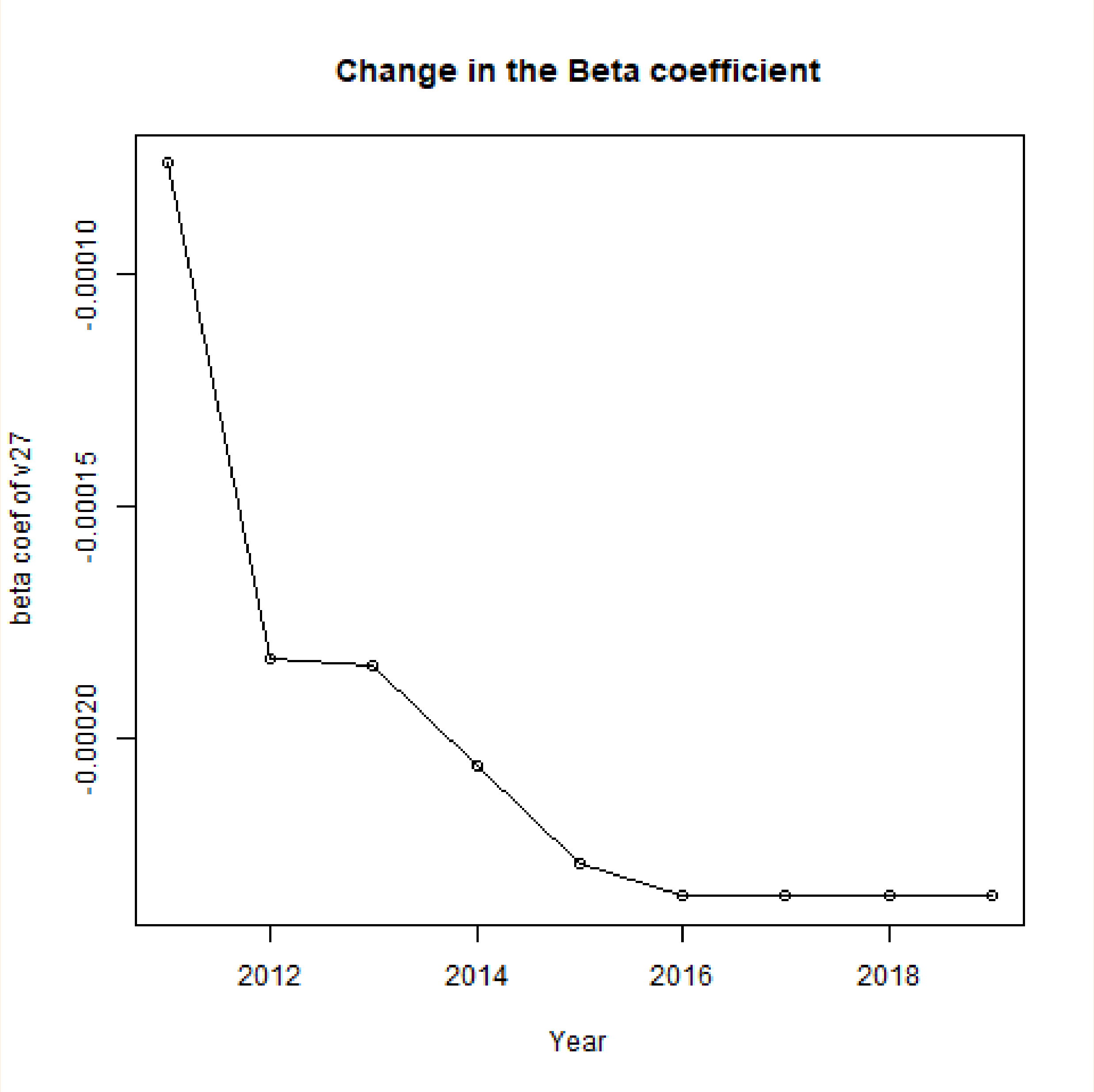
Impact of Asphyxia on Pneumonia

In the two plots, we can see trends in the value of the beta's over the years This could be due to many uncounted reasons.

Impact of having Newborn weight less than 2.5kg on Pneumonia



year wise



Accumulative wise

A similar change can be noticed in the beta's of the Newborn weight less than 2.5kg on Pneumonia

To find if there were any changes in the coefficients of the dependent Variables in Various Zones.

Steps Taken -

- Run Regression analysis on all six zones differently
- Get the coefficient and their standard error
- Then plot the confidence interval of these coefficients for all seven independent variables

State Classification by Geographical Zones:

North Zone: Himachal Pradesh, Punjab, Uttarakhand, Uttar Pradesh and Haryana.

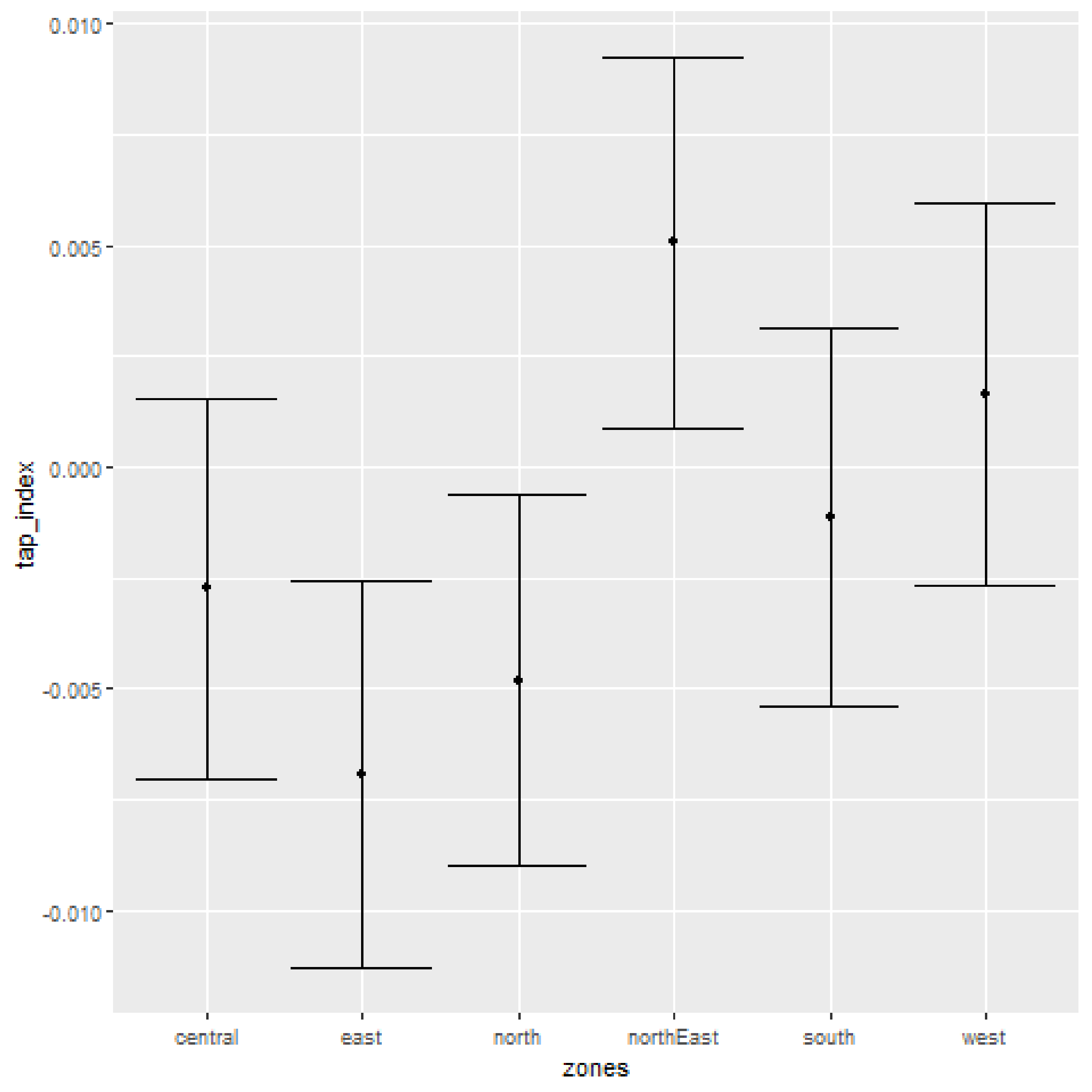
East Zone: Bihar, Orissa, Jharkhand, and West Bengal.

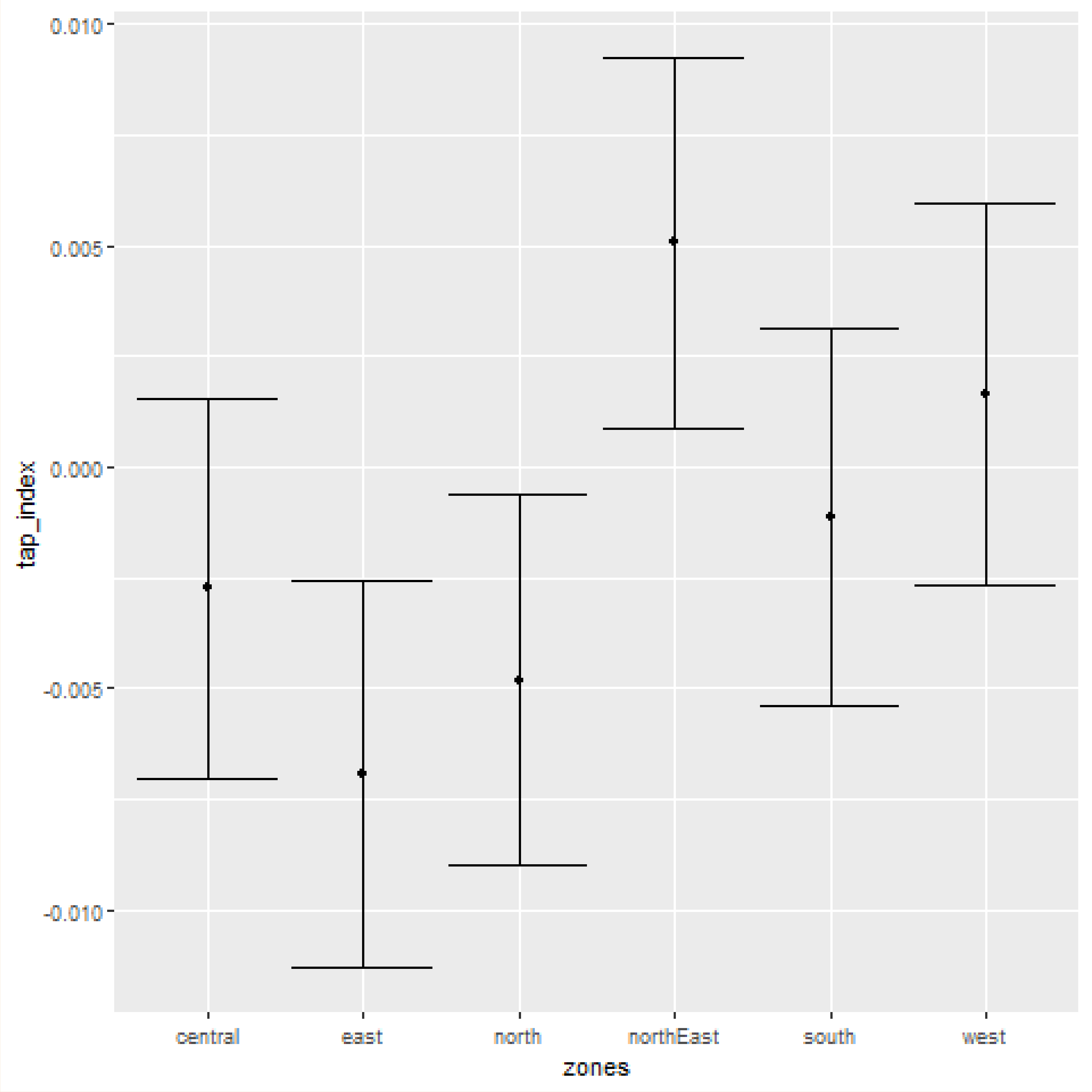
West Zone: Rajasthan, Gujarat, Goa and Maharashtra.

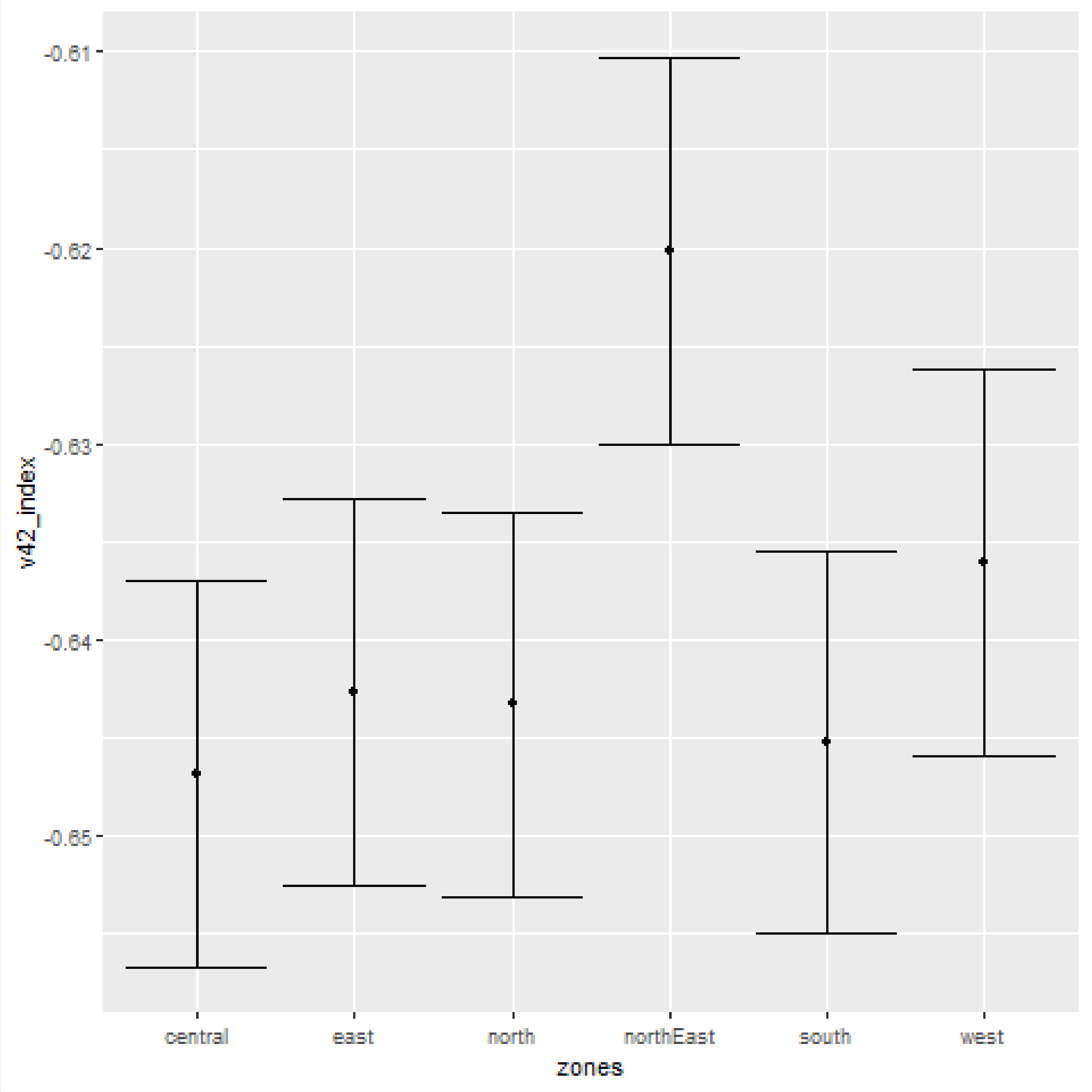
South Zone: Andhra Pradesh, Telangana, Karnataka, Kerala and Tamil Nadu

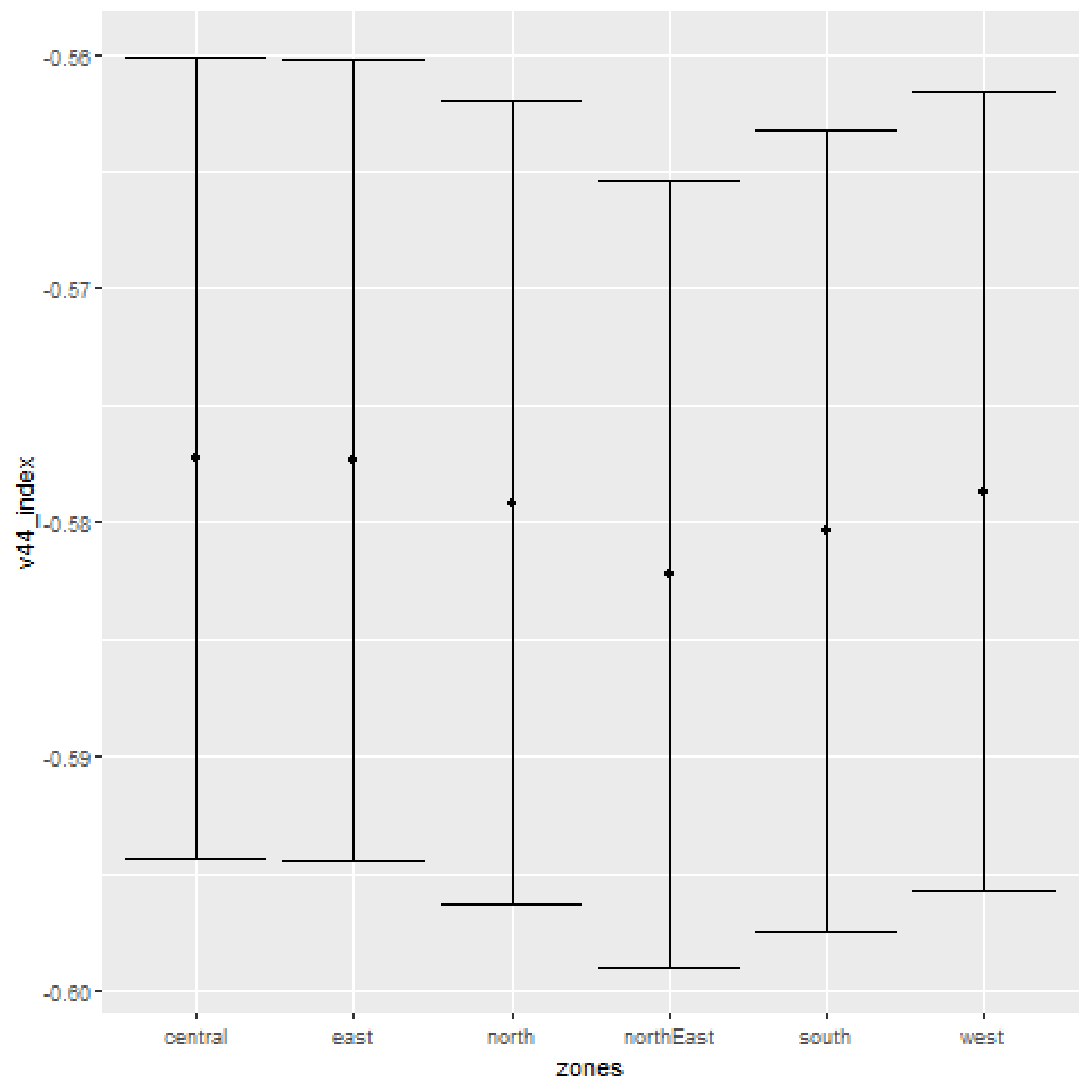
Central Zone: Madhya Pradesh and Chhattisgarh

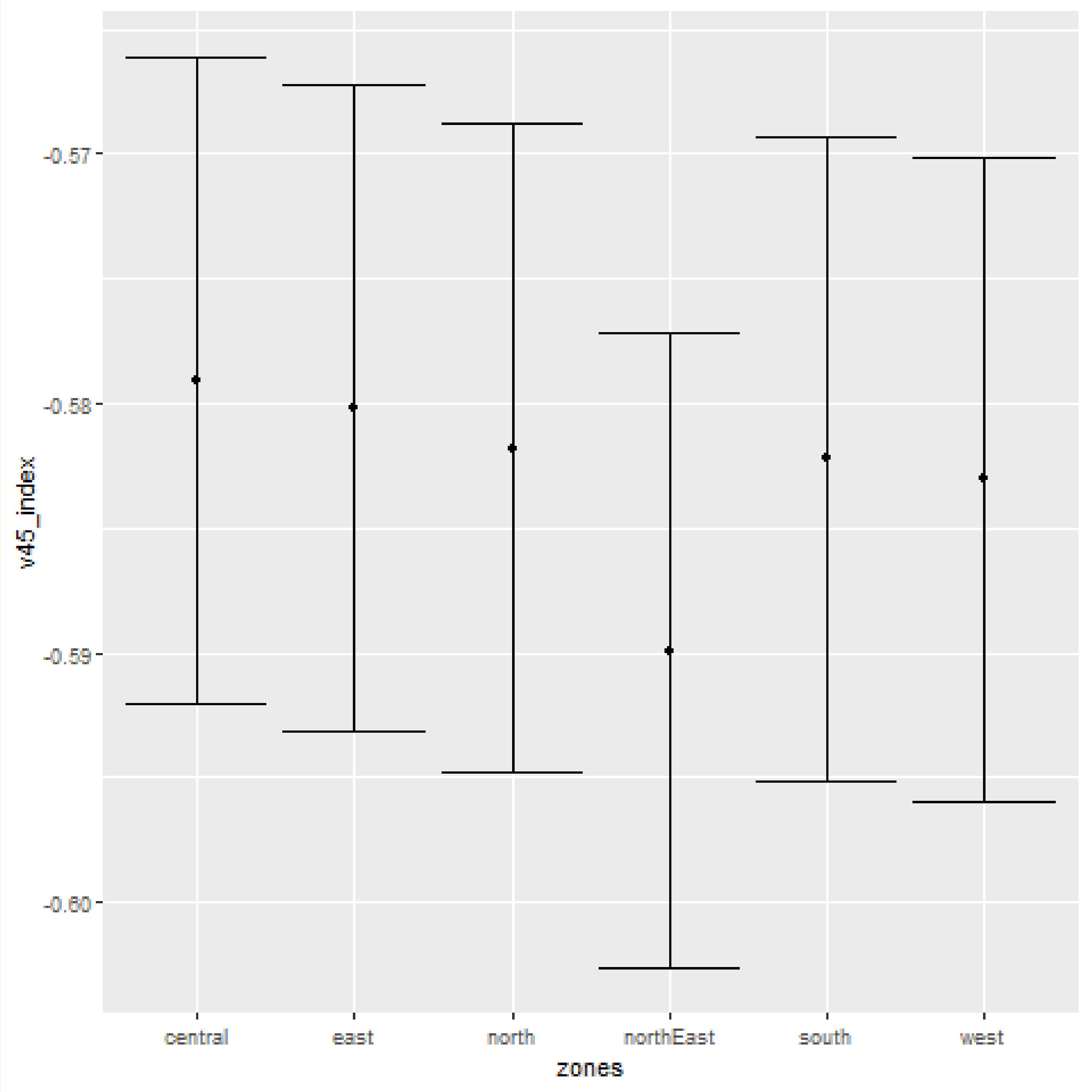
North East Zone: Assam, Sikkim, Nagaland, Meghalaya, Manipur, Mizoram, Tripura and Arunachal Pradesh

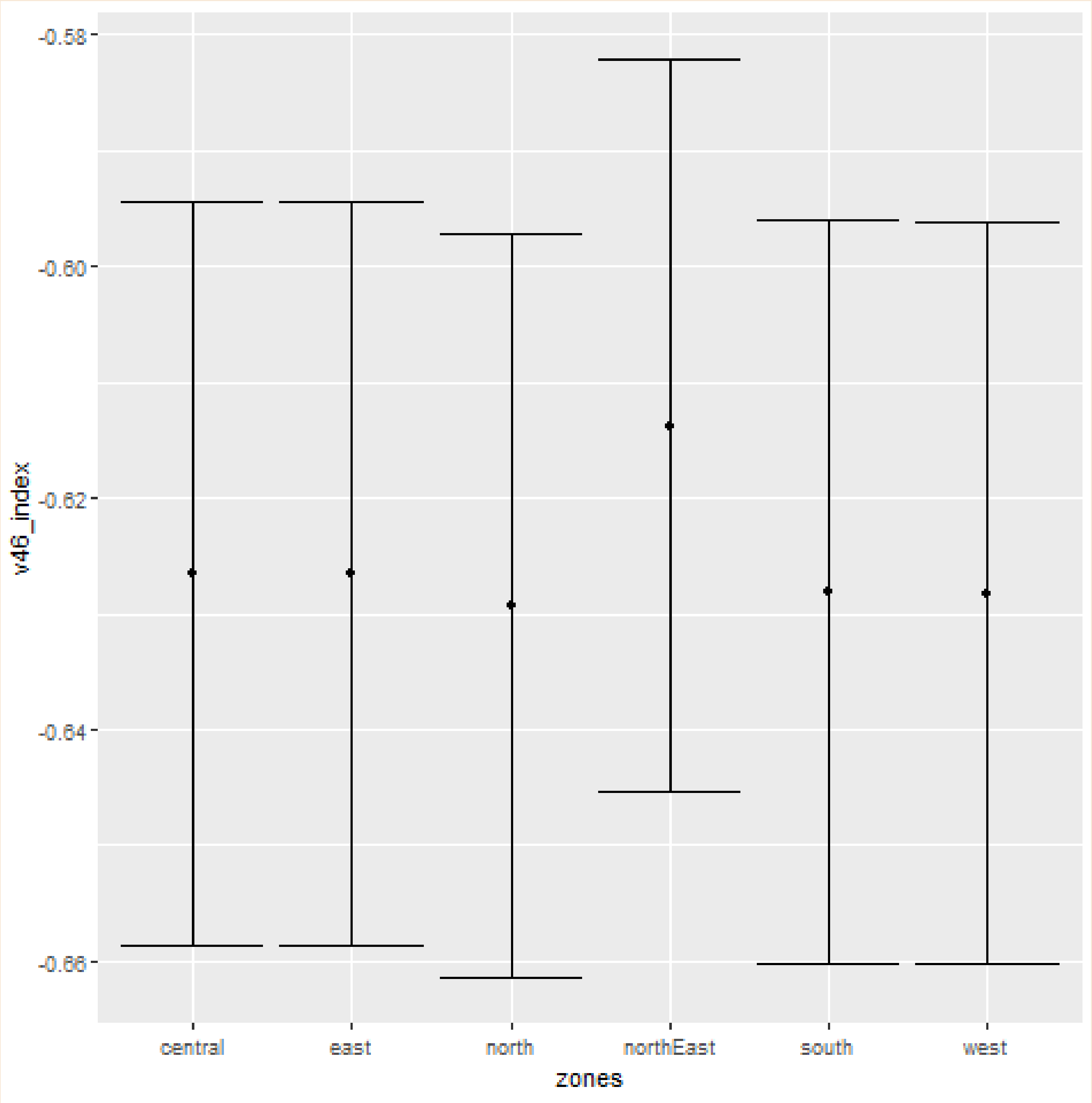


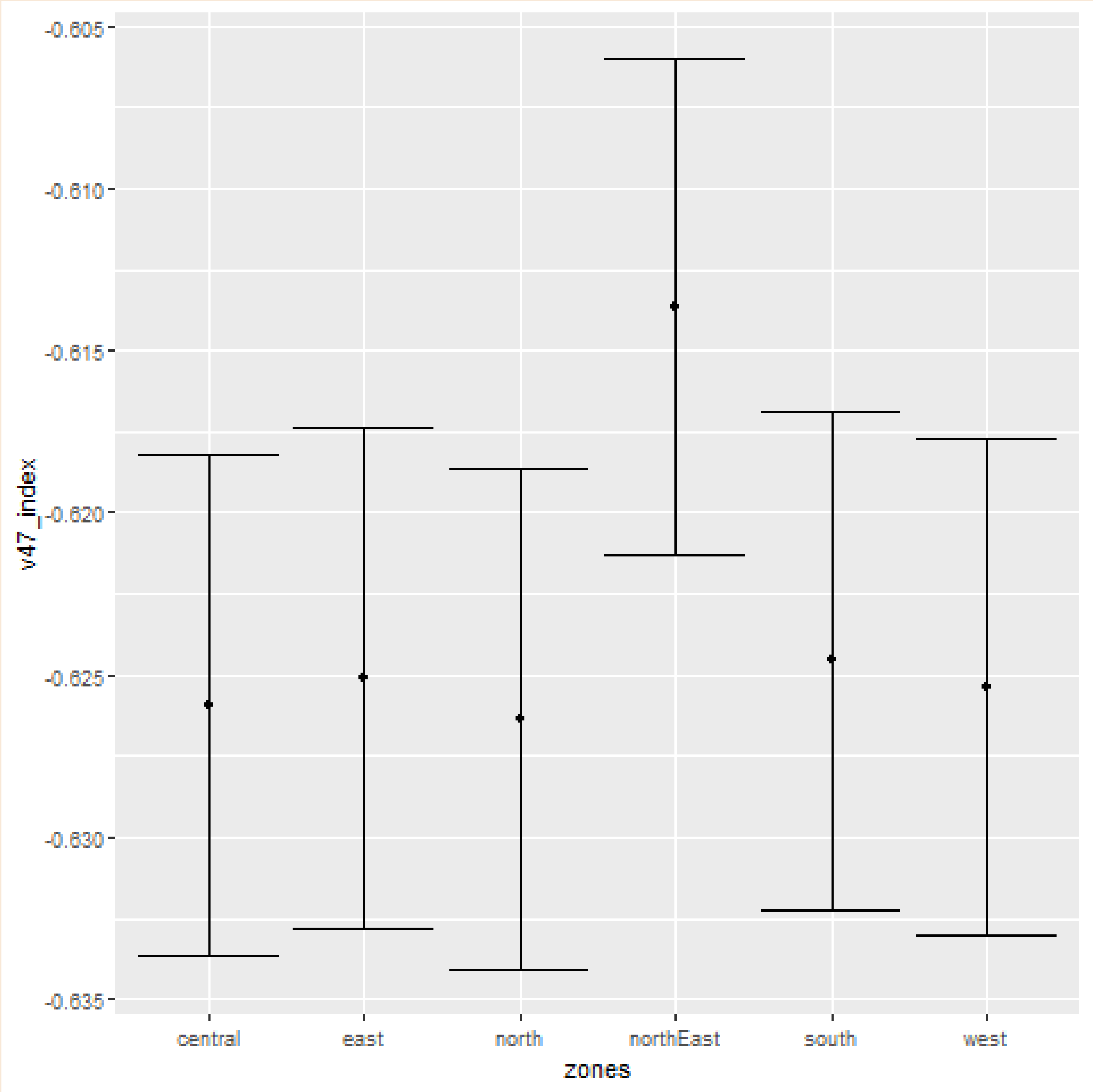

0.010
0.005
0.000
-0.005
-0.010
tap_index
central
east
north
northEast
south
west
zones


-0.61
-0.62
-0.63
-0.64
-0.65
v42_index
central
east
north
northEast
south
west
zones


-0.56
-0.57
-0.58
-0.59
-0.60
v44_index
central
east
north
northEast
south
west
zones


-0.57
-0.58
-0.59
-0.60
v45_index
central
east
north
northEast
south
west
zones





In these plots, we can notice that the coefficient of the independent variables for the northeast zone is different from the other zones. The reasons can be wet climate/ reachability or some unaccounted reasons.

TEAM MEMBERS

1. Mohit Sharma(2020086)
2. Aishwary Sharma(2020490)
3. Yash Agrawal (2020551)
4. Ujjwal Rastogi (2020546)
5. Mohit Bansal (2020526)

Thank you!