ECONOMETRICS-I

WINTER 2022

DATA ASSIGNMENT 2

Assigned Dependent Variable -

(v36) Percentage of children with Measles in the age group of 0 to 5 years

GROUP MEMBERS-

DIYA AHUJA (2020431)
AANYA TREHAN (2020419)
SHAGUN MOHTA (2020468)
PULKIT BAZARD (2020455)
AMOLIKA BANSAL (2020424)

Q1.

Our model

$$H_{i,t} = \beta_0 + \beta_1(log(gdp_{i,t})) + \beta_2(tap_{i,t}) + \beta_3(log(beds_{i,t})) + \beta_4(v45_{i,t}) + \beta_5(v37_{i,t}) + \beta_6(v16_{i,t}) + \beta_7(v21_{i,t}) + \beta_8(v28_{i,t}) + \beta_9(v46_{i,t}) + \beta_{10}(\log(v34_{i,t}))$$

Variable	Description	Reason for choosing
log(Gdp)	State Wise GDP	If the GDP of the state or district increases then generally it is assumed(because income inequality is not considered in GDP calculation) that the income of the citizens of that area also increases. Consequently increase in income leads to better access to healthcare and basic amenities of life and hence a lower chance of contracting measles.
tap	District Wise Tap Water Access (Percentage of Households) as of 2019	Measles is a highly contagious disease and hence can be easily spread among many people. Still, a huge part of India does not have access to safe and clean drinking water, which can prove to be very dangerous in case of a measles outbreak.
log(beds)	State Wise Number of Hospital Beds (as of 2020)	In case of a measles outbreak, the number of beds in hospitals will increase to accommodate the rise in demand for hospital beds, similar to what had happened in covid as both are highly transmissible diseases.
v45	Percentage of infant deaths due to Fever (to total reported infant deaths)	Fever is one of the symptoms of measles and can run very high in the patients leading to severe consequences and potentially leading to death due to complications. The early signs of measles start with fever, and in case people are unaware of the symptoms associated with measles, it can lead to an unidentified death, especially in areas of low literacy and awareness.
v37	Percentage of children with Diarrhea and Dehydration in the age group of 0 to 5 years	Diarrhea and dehydration are a few of the symptoms of measles and are generally categorized as complications associated with the disease, specifically in children under the age of 5 years. Thus, the percentage of children with measles is correlated with the percentage of children with diarrhea and dehydration.
v16	Percentage of safe deliveries (to total reported deliveries)	An unsafe delivery typically depicts a weaker immunity in the newborn and a higher chance of

		contracting infections such as measles. The complications can also be expected to be much worse in the case of a newborn with unsafe delivery.
v21	Percentage of women received a postpartum checkup or Post-Natal Care between 48 hours to 14 days of delivery.	After birth, women and infants require assistance and close supervision. Mothers usually take their newborns along with them to their postpartum checkups. According to WHO, the majority of mother and baby fatalities occur within the first six weeks following delivery.
v28	Percentage of newborns having weight less than 2.5 kg	Children born with low birth weight, that is less than 2.5 kg are at a significant risk of developing diseases like measles in upcoming years
v46	Percentage of infant deaths due to Measles (to total reported infant deaths)	Measles can be fatal in people of all ages. However, complications are more likely in children under the age of five and adults over the age of twenty. Ear infections and diarrhea are common complications. Pneumonia and encephalitis are serious complications.
log(v34)	Fully immunized children in the age group of 9 to 11 months	It has been observed that immunization has caused a drastic decline in measles. Fully immunized children in the age group of 9 to 11 months are at a much lower risk of contracting measles and hence affect the total percentage of children with measles in the age group pf 0 to 5 years.

The main reasons for choosing the above model is based on factors

1. The adjusted R^2 value or the Goodness of fit for our model came out to be feasible.Only when a new variable improves the model more than by chance does adjusted R-squared increase. Variables of poor quality can cause it to fall. Based on these observations, we chose the explanatory variables.
2. The variables chosen are relevant and align well theoretically with our response variable.
3. To improve our previous model, with no log terms, we transformed some variables, using a “log” transformation.After transforming the variables gdp,beds and v34 , our model’s distribution, regression r-squared, and residual plot patterns changed. These variables were not provided in percentage form and hence we took log for scalability so that our model gives more accurate estimations.
   Before doing log transformation for the Kharif data:
   Adjusted R-squared= 0.09184
   sum(residual error) = 1.17373e-10
   After doing log transformation for the Kharif data:
   Adjusted R-squared= 0.403

sum(residual error) =-3.739896e-11(much smaller and closer to 0)

4. The residuals in our model were evenly distributed and have an average value of approx zero

```
main <- do.call(data.frame,lapply(main,function(x) replace(x, is.infinite(x), NA)))
kharif_data <- subset(main,main$season=="Kharif")
kharif1 <- do.call(data.frame,lapply(kharif_data,function(x) replace(x, is.infinite(x), NA)))
rabi_data <- subset(main,main$season=="Rabi")
main$lgdp <- log(main$gdp)
main$lv34 <- log(main$v34)
main$lbeds <- log(main$beds)
model1 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 + lv34,data =
kharif_data)
summary(model1)
model2 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 + lv34,data =
rabi_data)
summary(model2)
```

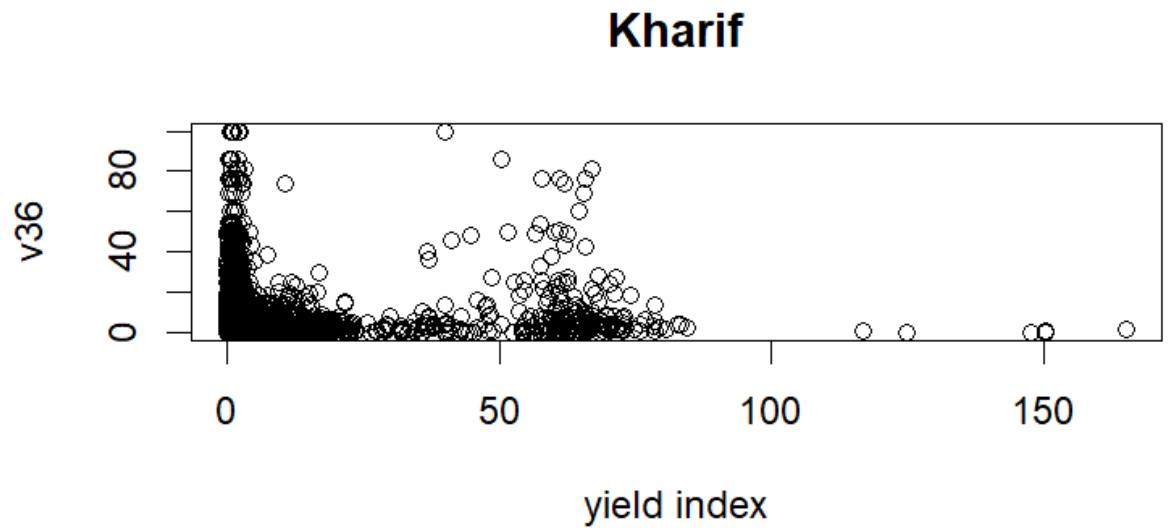
```
> model1 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46
 + v45 + lv34,data = kharif_data)
> summary(model1)

Call:
lm(formula = v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 +
    v46 + v45 + lv34, data = kharif_data)

Residuals:
    Min      1Q  Median      3Q     Max
-21.900  -1.819  -0.362   1.298  57.868

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 17.096971   0.670163  25.512  < 2e-16 ***
lgdp        -1.267669   0.065318 -19.408  < 2e-16 ***
tap         -0.003432   0.001757  -1.953 0.050819 .
lbeds        1.962055   0.070582  27.798  < 2e-16 ***
v37         -0.197232   0.002613 -75.484  < 2e-16 ***
v16          0.028951   0.003789   7.640 2.29e-14 ***
v21         -0.007812   0.001273  -6.135 8.72e-10 ***
v28         -0.057706   0.004474 -12.899  < 2e-16 ***
v46          0.588727   0.012397  47.490  < 2e-16 ***
v45          0.102262   0.004238  24.129  < 2e-16 ***
lv34         0.193297   0.052721   3.666 0.000247 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.187 on 17233 degrees of freedom
  (11585 observations deleted due to missingness)
Multiple R-squared:  0.4053,    Adjusted R-squared:  0.405
```

```
> model2 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46
 + v45 + lv34,data = rabi_data)
> summary(model2)

Call:
lm(formula = v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 +
    v46 + v45 + lv34, data = rabi_data)

Residuals:
    Min      1Q  Median      3Q     Max
-20.840  -1.938  -0.378   1.283  58.667

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.485390   0.707064  20.487  < 2e-16 ***
lgdp        -1.308331   0.069113 -18.930  < 2e-16 ***
tap         -0.003428   0.001910  -1.795   0.0728 .
lbeds        2.056134   0.073983  27.792  < 2e-16 ***
v37         -0.183441   0.002649 -69.236  < 2e-16 ***
v16          0.030586   0.003946   7.752 9.62e-15 ***
v21         -0.007861   0.001356  -5.797 6.87e-09 ***
v28         -0.066459   0.005027 -13.220  < 2e-16 ***
v46          0.595936   0.013766  43.290  < 2e-16 ***
v45          0.106447   0.004529  23.502  < 2e-16 ***
lv34         0.286136   0.057823   4.948 7.56e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.29 on 15393 degrees of freedom
  (10334 observations deleted due to missingness)
Multiple R-squared:  0.3906,    Adjusted R-squared:  0.3902
```

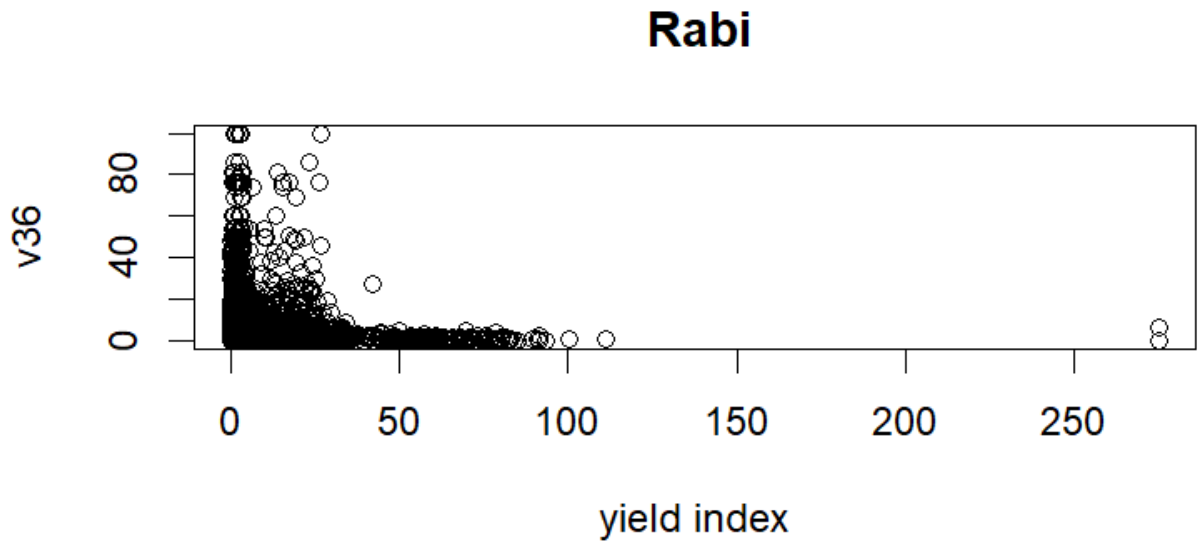
(b)

```
kharif2<-subset(kharif_data,v36!="NA")
kharif2<-subset(kharif2,v37!="NA")
kharif2<-subset(kharif2,v16!="NA")
kharif2<-subset(kharif2,v21!="NA")
kharif2<-subset(kharif2,v46!="NA")
kharif2<-subset(kharif2,v45!="NA")
kharif2<-subset(kharif2,tap!="NA")
kharif2<-subset(kharif2,beds!="NA")
kharif2<-subset(kharif2,v34!="NA")
kharif2<-subset(kharif2,gdp!="NA")
kharif2<-subset(kharif2,v28!="NA")

rabi2<-subset(rabi_data,v36!="NA")
rabi2<-subset(rabi2,v37!="NA")
rabi2<-subset(rabi2,v16!="NA")
rabi2<-subset(rabi2,v21!="NA")
rabi2<-subset(rabi2,v46!="NA")
rabi2<-subset(rabi2,v45!="NA")
rabi2<-subset(rabi2,tap!="NA")
rabi2<-subset(rabi2,beds!="NA")
rabi2<-subset(rabi2,v34!="NA")
rabi2<-subset(rabi2,gdp!="NA")
rabi2<-subset(rabi2,v28!="NA")
```

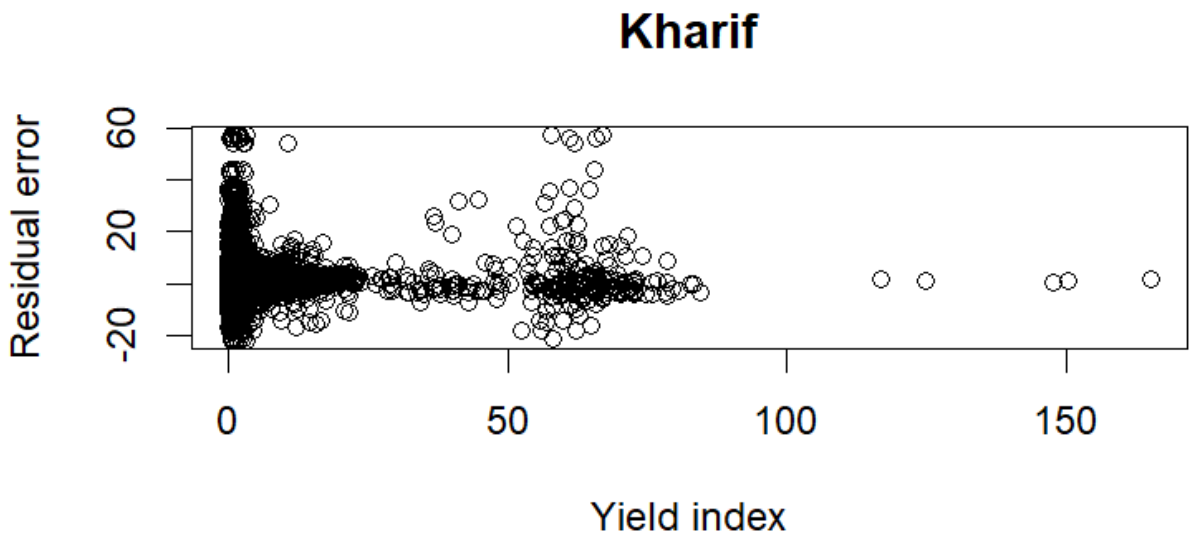
```
plot(kharif2$index,  kharif2$v36, ylab="v36", xlab="yield index", main="Kharif")
```



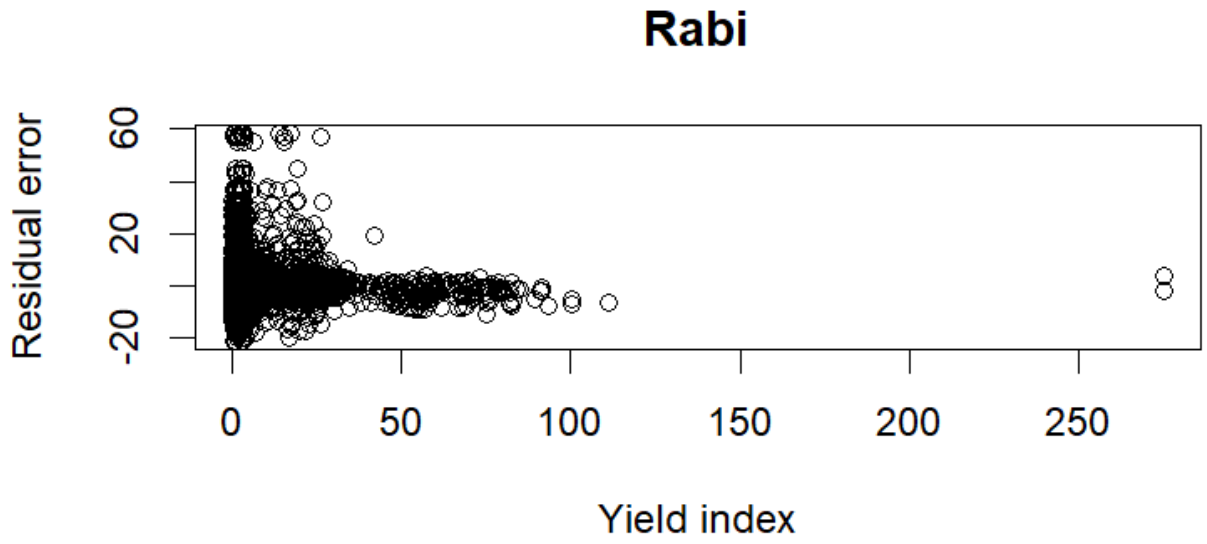

```
plot(rabi2$index,  rabi2$v36, ylab="v36", xlab="yield index", main=" Rabi")
```



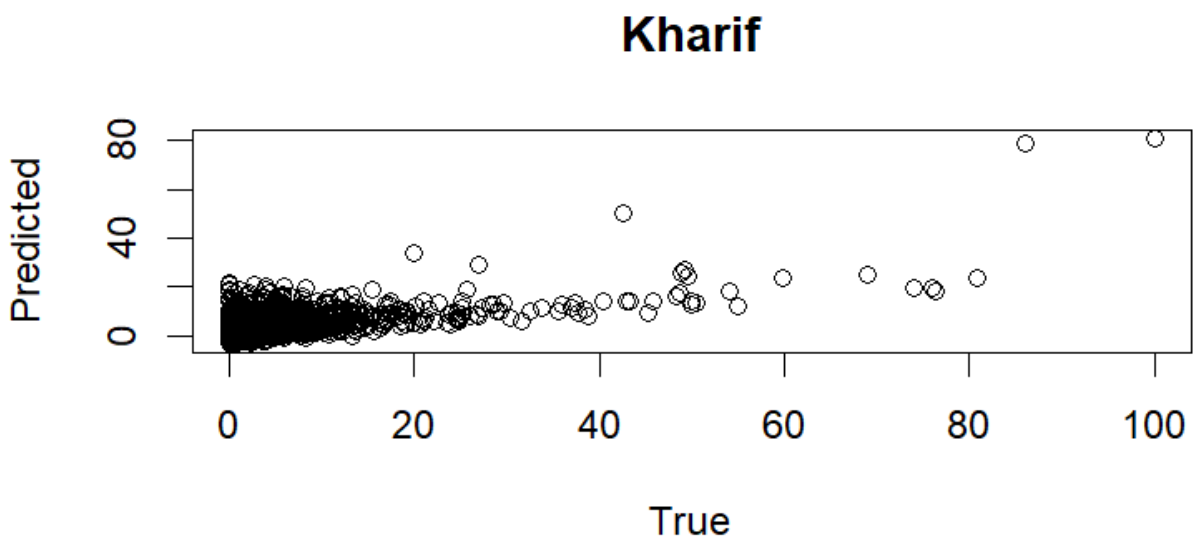
```
# calculating the residual value for kharif
res1 = resid(model1)
# ûi,t on y-axis and yield index on x-axis for kharif.
plot(kharif2$index,res1, ylab="Residual error", xlab="Yield index", main="Kharif")
```



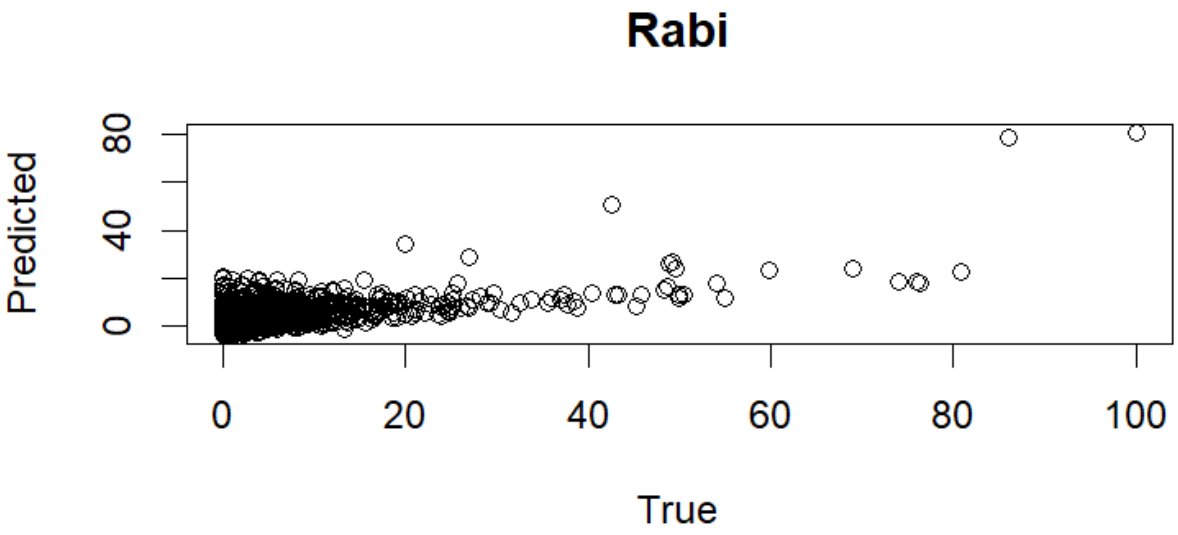

```
# calculating the residual value for Rabi
res2 = resid(model2)
# ûi,t on y-axis and yield index on x-axis for Rabi.
plot(rabi2$index,res2, ylab="Residual error", xlab="Yield index", main="Rabi")
```



```
#predicted values of the health indicator on y-axis and true values of the health indictor on x-axis Kharif
plot(kharif2$v36,predict(model1), ylab="Predicted", xlab="True", main="Kharif")
```




```
#predicted values of the health indicator on y-axis and true values of the health indictor on x-axis Rabi
plot(rabi2$v36,predict(model2), ylab="Predicted", xlab="True", main="Rabi")
```



Interpretation of all the graphs

The residuals are the difference between predicted values by our model and the actual/true values in the recorded dataset**.**
Residual=Observed-Predicted
Thus predicted=Observed/true-Residual

Using the graph points in the third graph, where we have established a plot between observed and predicted values, we can find the residual values for each observation v36.

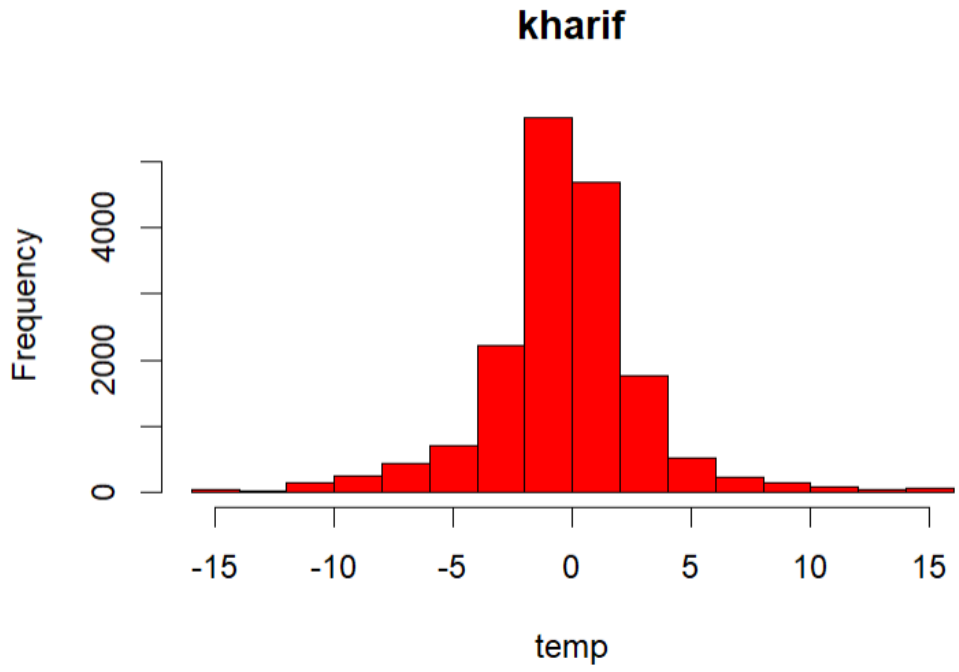
Similarly using the plot between v36 and yield index in the first plot, therefore we can easily extract the value of yield index for a particular observation of v36.

Now, the yield index and residual values obtained from the first and third plots match with the values from the second plot, proving the relationship between residual error, predicted, and true/observed value.

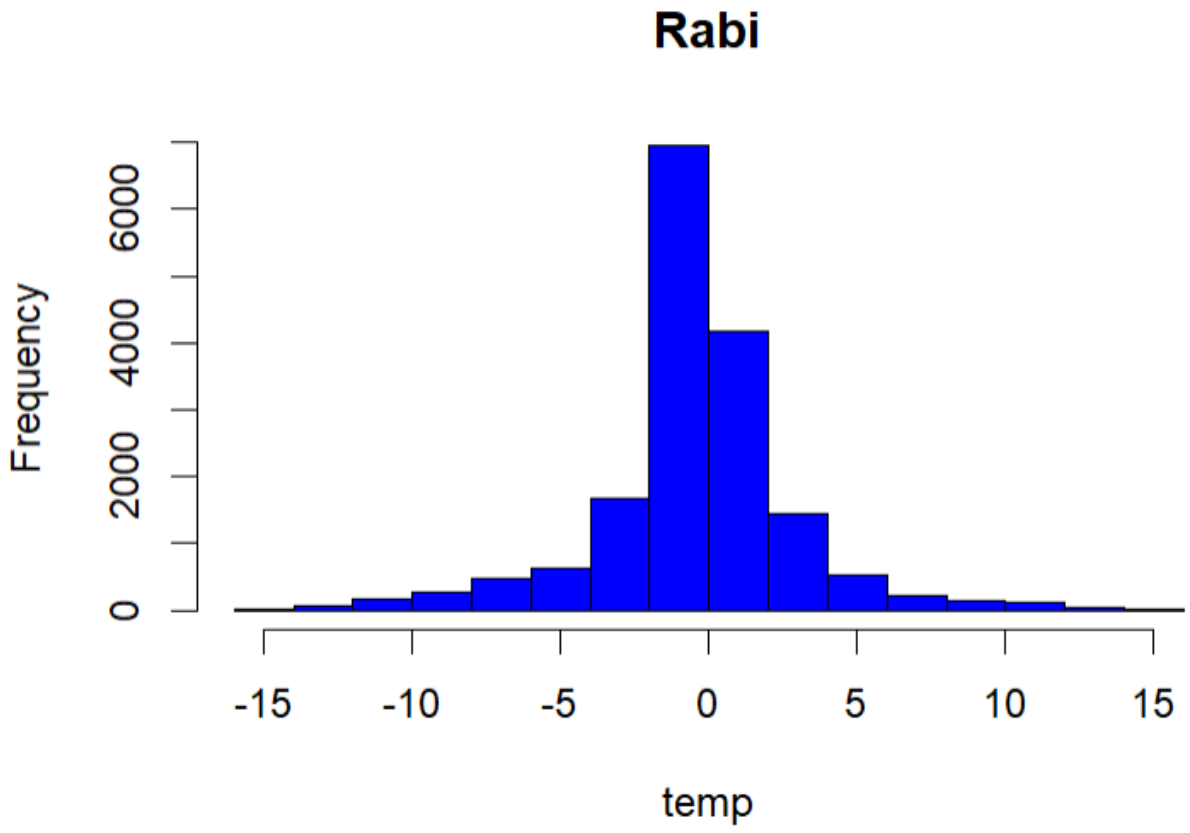
(c)

```
> #Plot a histogram of ûi,t and verify that Σi,t ûi,t = 0
> # for kharif
>
> x<- res1
> temp<- res1[x>- (mean(res1)+ 3*sd(res1)) & x< mean(res1)+ 3
*sd(res1)]
>
> hist(temp, col='red', main='kharif')
>
>
> sum(resid(model1))
[1] -1.30555e-11
```

The total of all differences between our fitted values (on the regression line) and the actual values above the line equals the sum of all differences between the regression line and all values below the line. Since the sum of all the residual values will be the sum of equal positive and negative values the value of sum(residual) comes out to be very close to 0 or approximately 0.



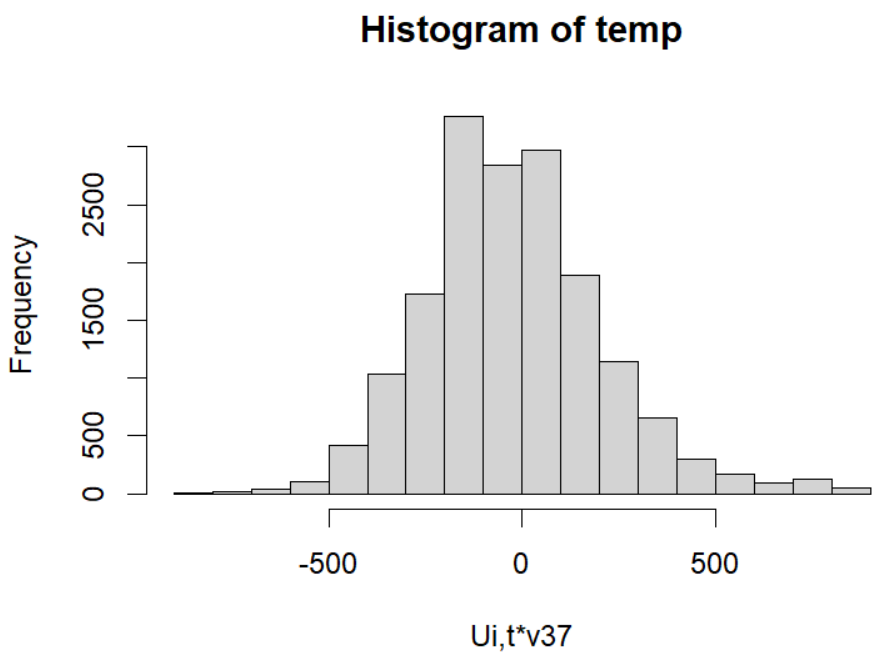
```
> #Plot a histogram of ûi,t and verify that Σi,t ûi,t = 0
> # for rabi
>
> x<- res2
> temp<- res2[x>- (mean(res1)+ 3*sd(res2)) & x< mean(res2)+ 3
*sd(res2)]
>
> hist(temp, col='blue', main='Rabi')
>
> sum(resid(model2))
[1] -1.822106e-11
>
```



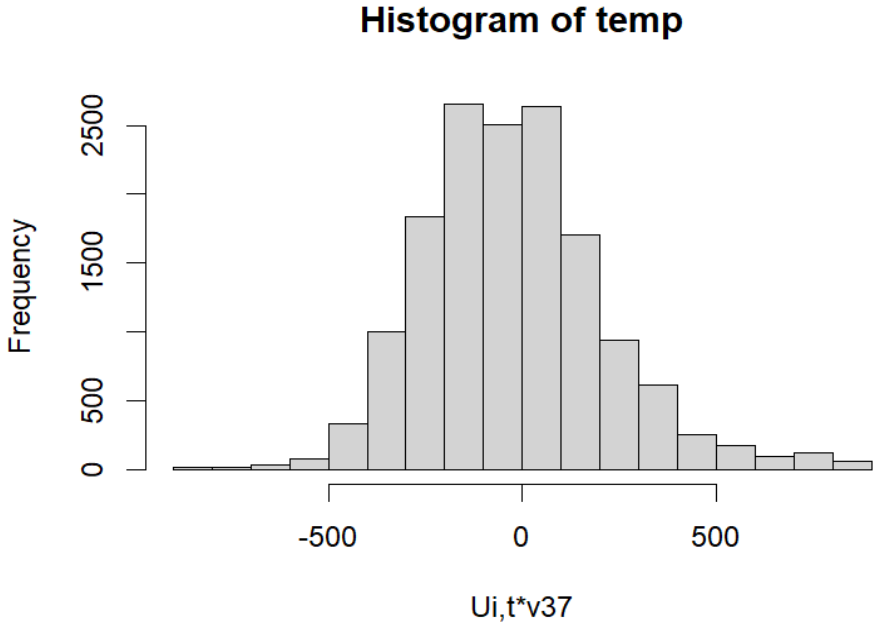
The magnitude of error terms is summarized in the histogram. We can observe how frequently errors occurred and it provides information about the bandwidth of errors.

1 d)

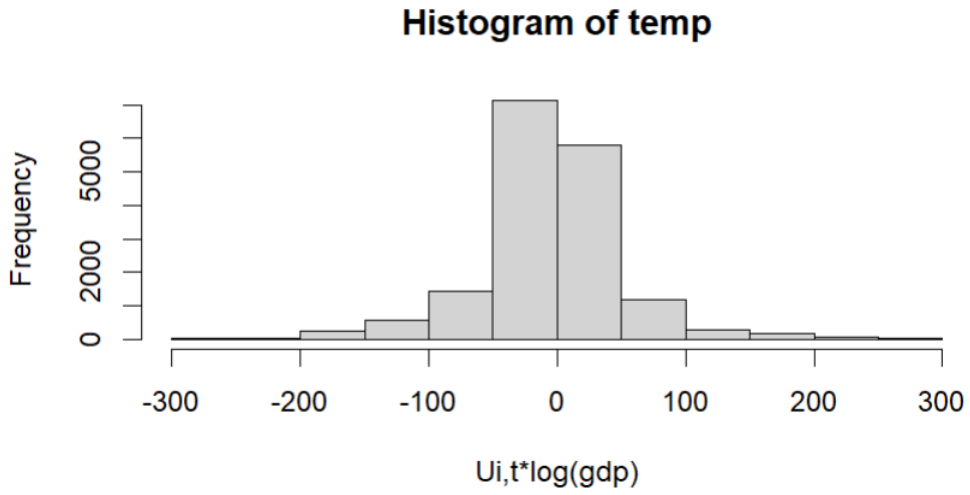
```
sum(temp<-res1*kharif2$v37)
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*v37")
```



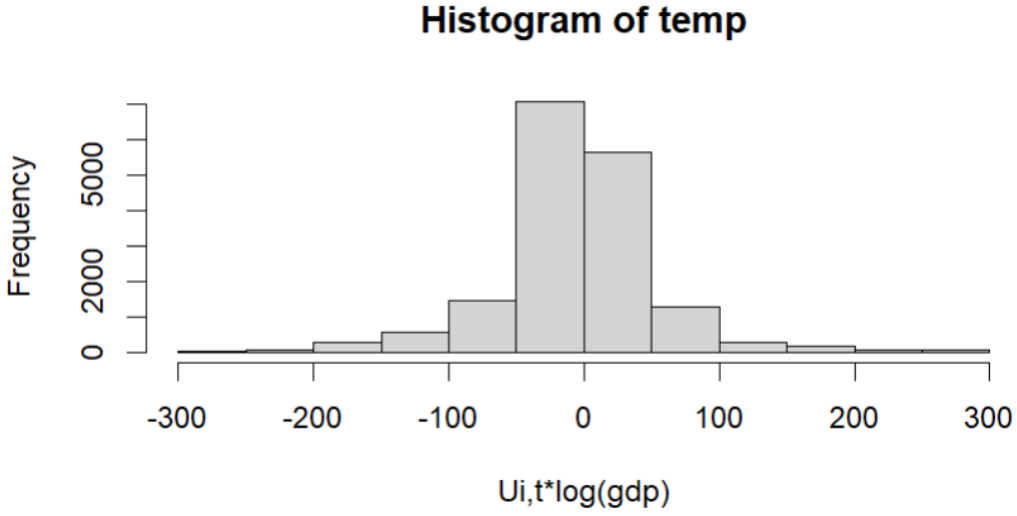

```
sum(temp<-res2*rabi2$v37)
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*v37")
```



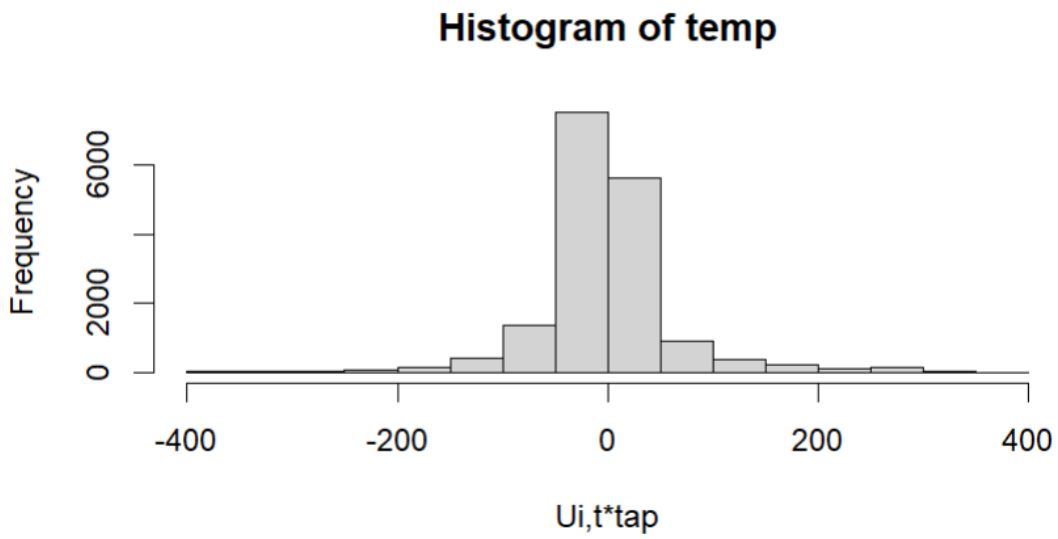
```
sum(temp<-res1*log(kharif2$gdp))
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*log(gdp)")
```



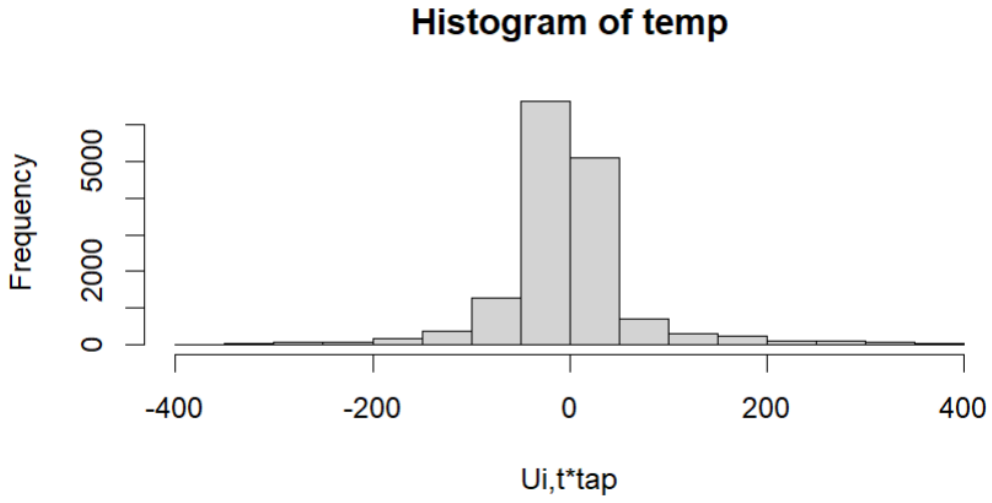

```
sum(temp<-res1*log(rabi2$gdp))
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*log(gdp)")
```



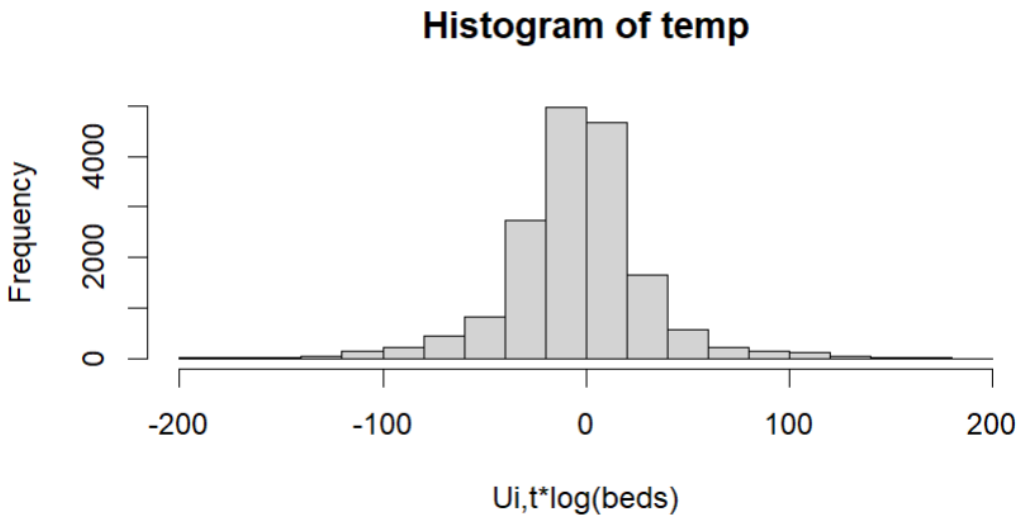
```
sum(temp<-res1*kharif2$tap)
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*tap")
```



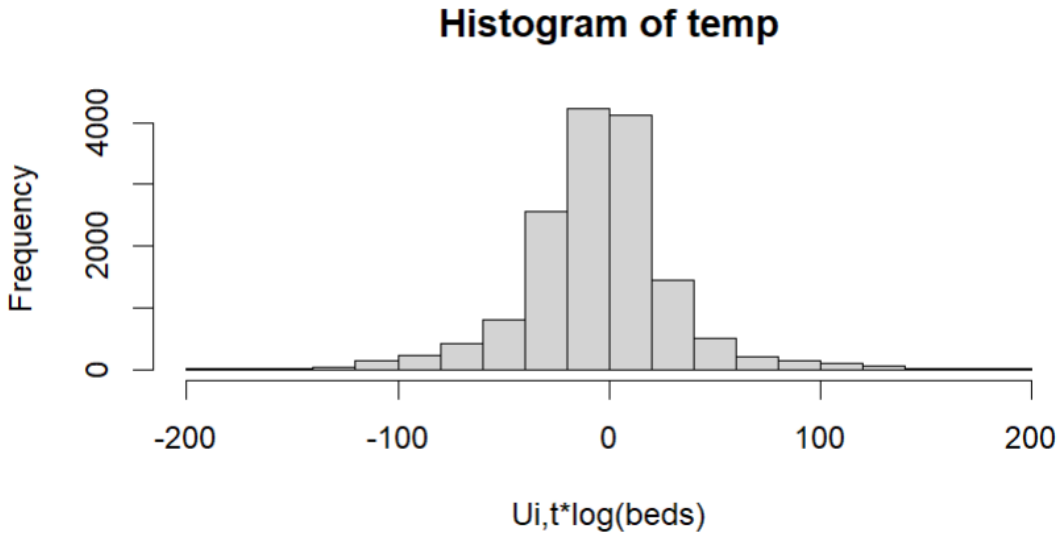

```
sum(temp<-res2*rabi2$tap)
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*tap")
```



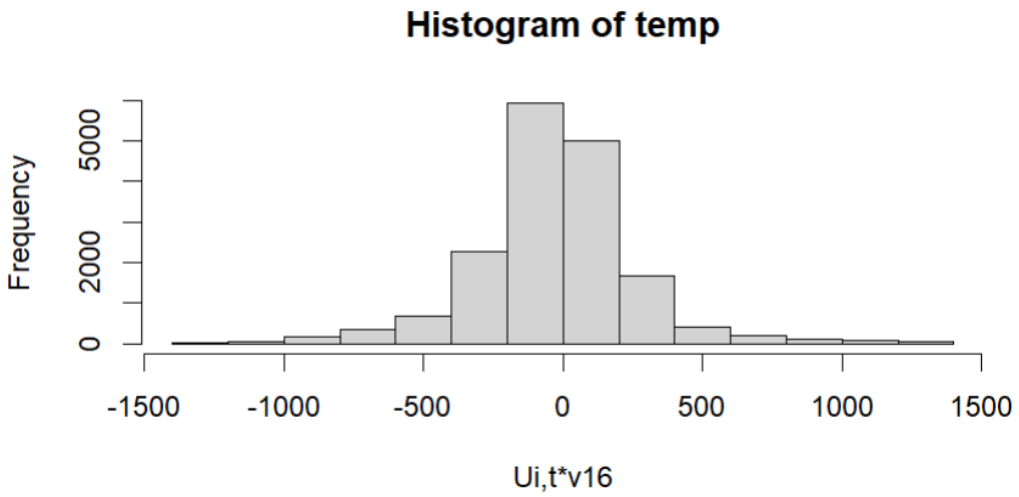
```
sum(temp<-res1*log(kharif2$beds))
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*log(beds)")
```



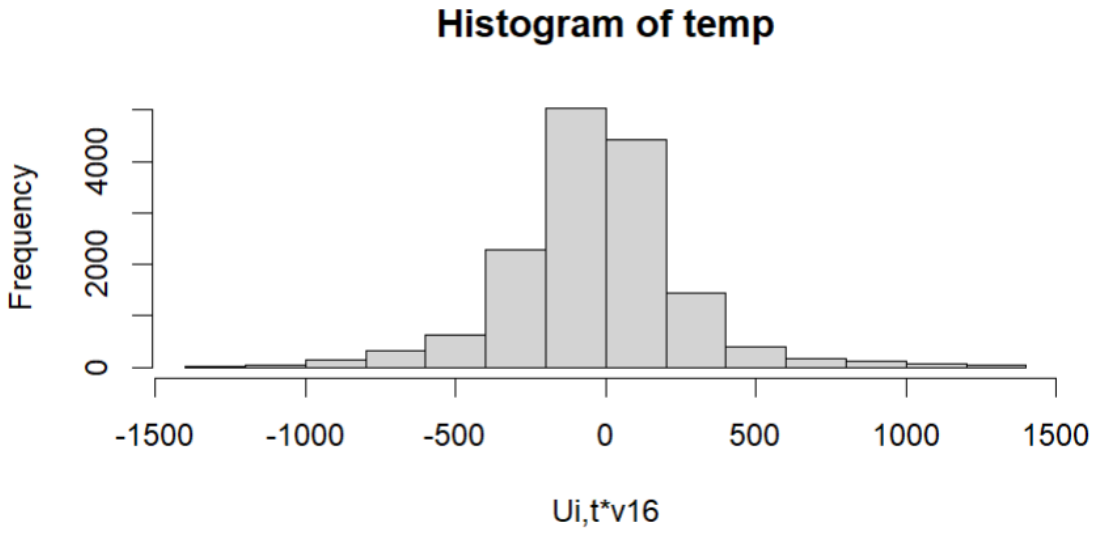

```
sum(temp<-res2*log(rabi2$beds))
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*log(beds)")
```



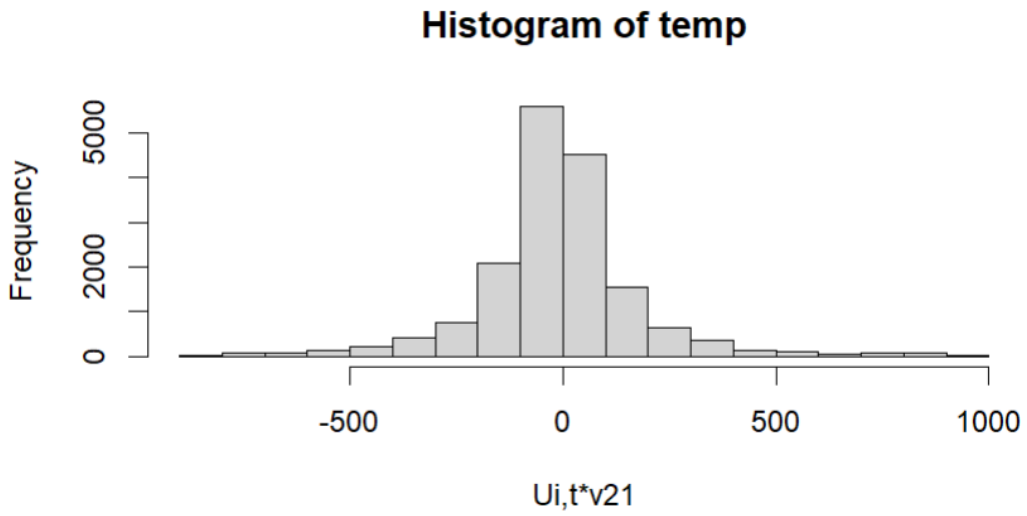
```
sum(temp<-res1*kharif2$v16)
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*v16")
```



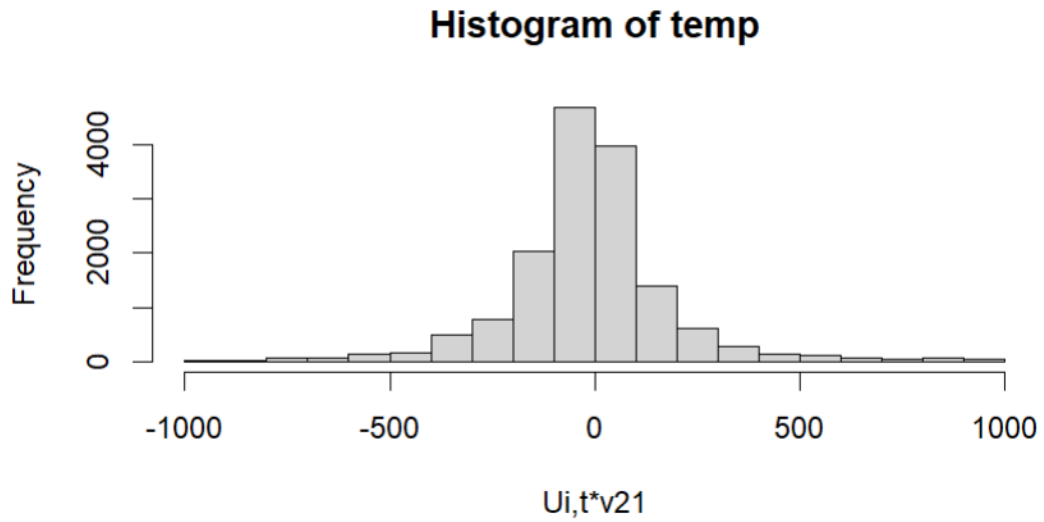

```
sum(temp<-res2*rabi2$v16)
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*v16")
```



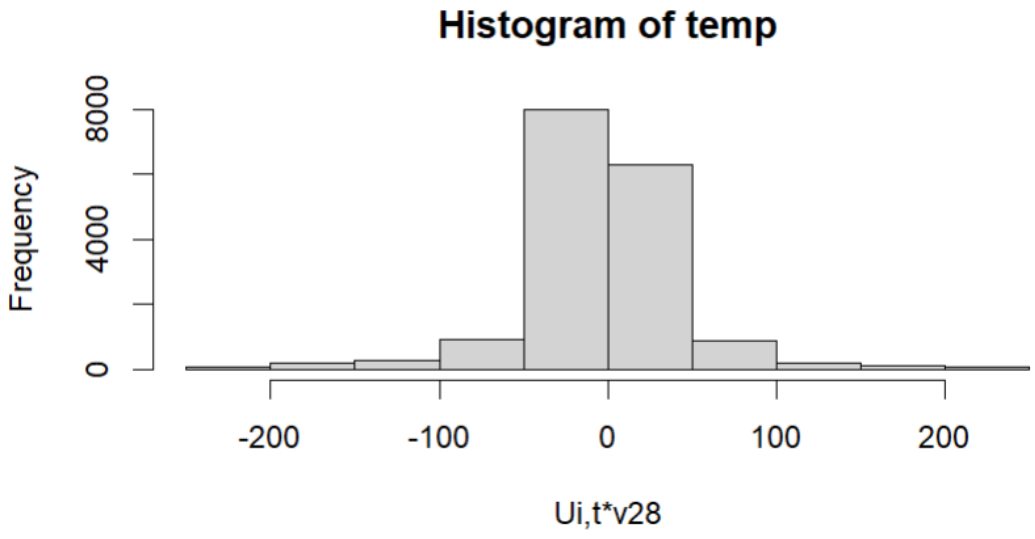
```
sum(temp<-res1*kharif2$v21)
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*v21")
```



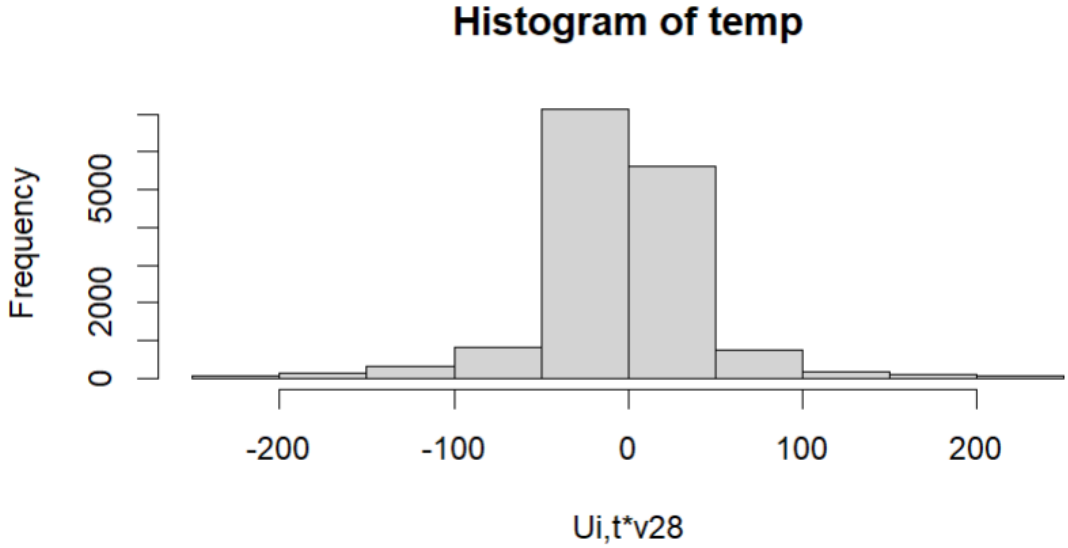

```
sum(temp<-res2*rabi2$v21)
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*v21")
```



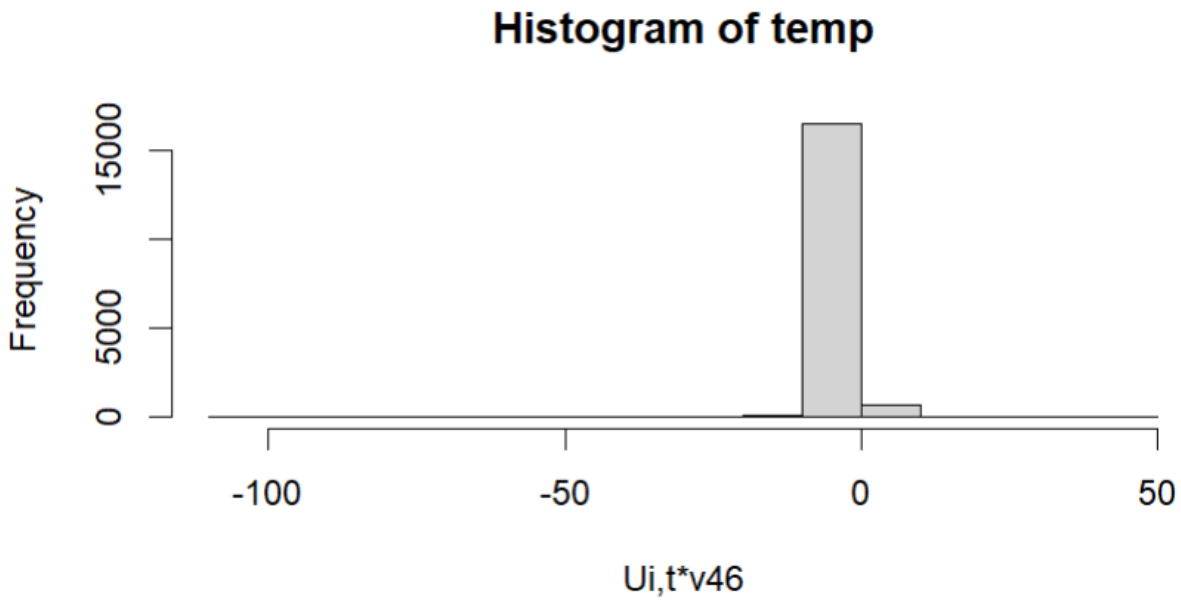
```
sum(temp<-res1*kharif2$v28)
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*v28")
```



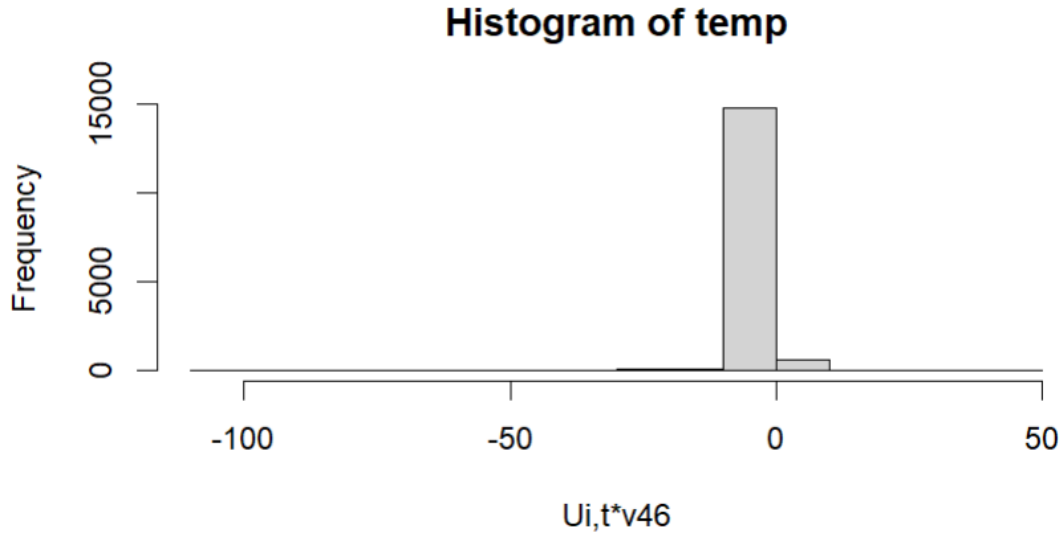

```
sum(temp<-res2*rabi2$v28)
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*v28")
```



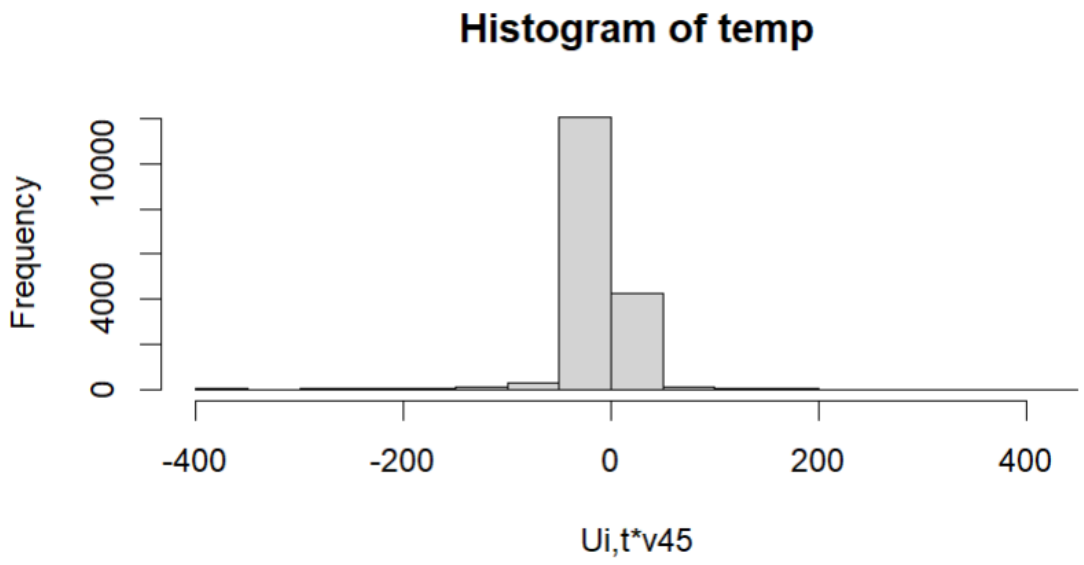
```
sum(temp<-res1*kharif2$v46)
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*v46")
```



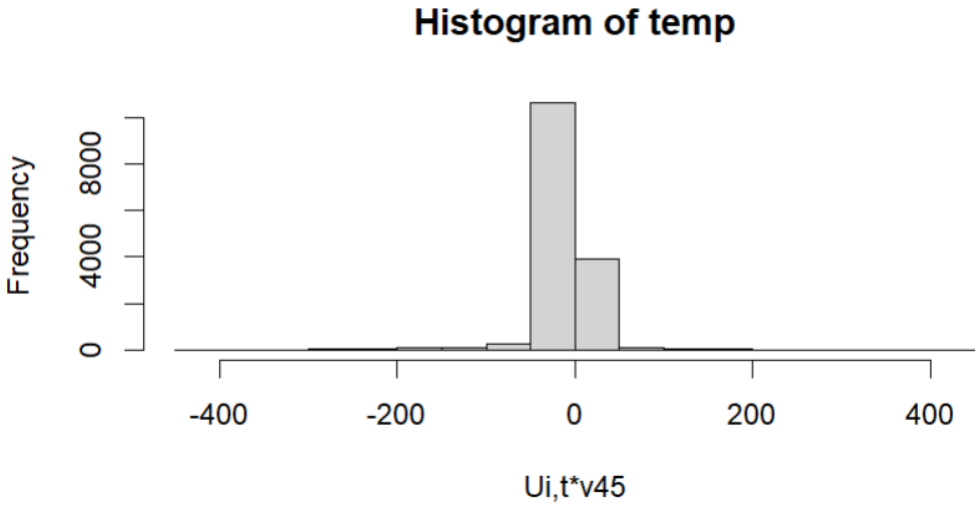

```
sum(temp<-res2*rabi2$v46)
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*v46")
```



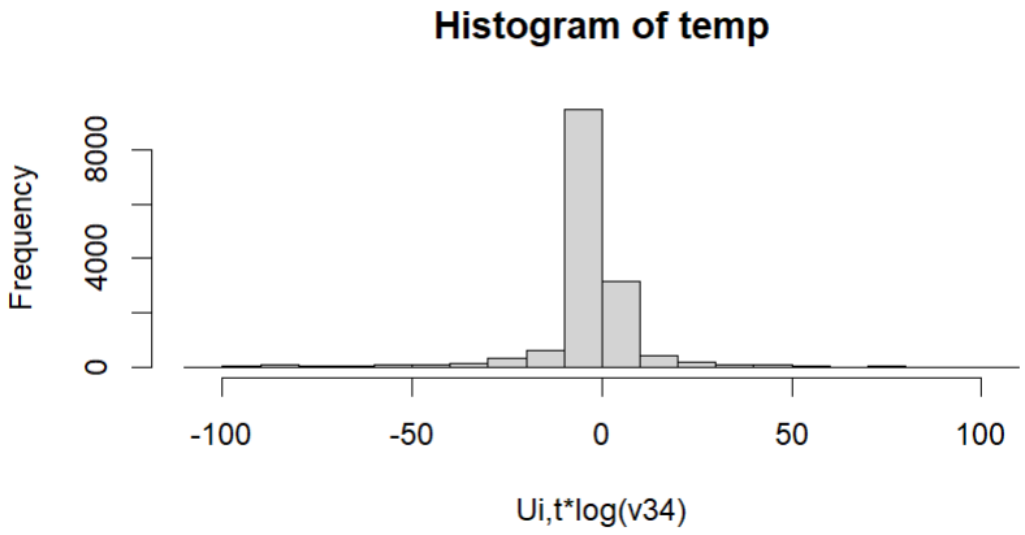
```
sum(temp<-res1*kharif2$v45)
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*v45")
```



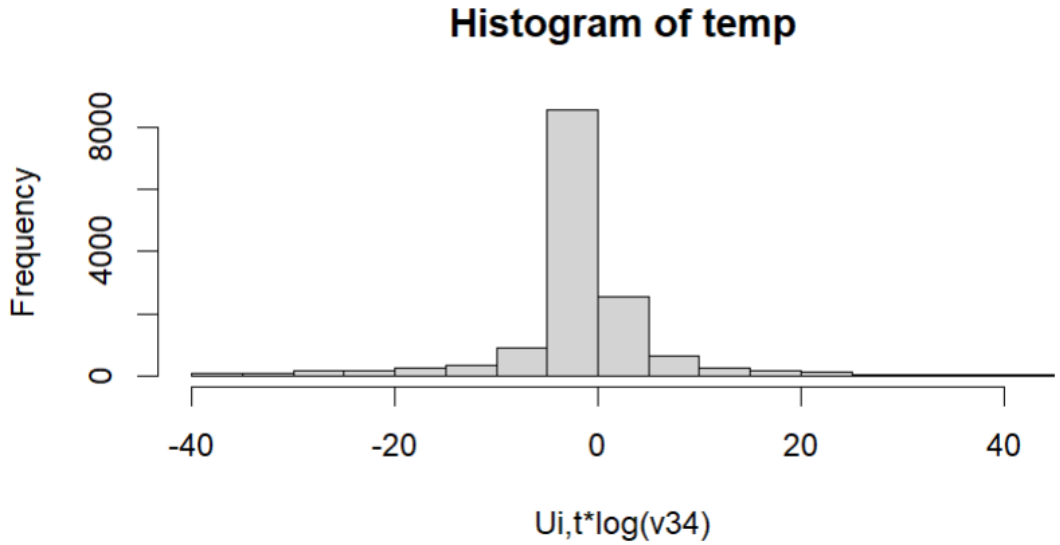

```
sum(temp<-res2*rabi2$v45)
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*v45")
```



```
sum(temp<-res1*kharif2$log(v34))
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*log(v34)")
```




```
sum(temp<-res2*rabi2$log(v34))
temp<- temp[temp>-(mean(temp)+3*sd(temp)) & temp<(mean(temp)+3*sd(temp)) ]
hist(temp, xlab = "Ui,t*log(v34)")
```



*Note -In the **above histograms**, the outliers have been reduced explicitly to produce a better fitted histogram.*

```
> sum(temp<-res1*kharif2$v37)
[1] 3.51997e-09
> sum(temp<-res1*log(kharif2$gdp))
[1] 5.804981e-10
> sum(temp<-res1*kharif2$tap)
[1] 3.528292e-10
> sum(temp<-res1*log(kharif2$beds))
[1] 5.72268e-10
> sum(temp<-res1*kharif2$v16)
[1] 3.743487e-09
> sum(temp<-res1*kharif2$v21)
[1] -6.772911e-10
> sum(temp<-res1*kharif2$v28)
[1] -3.656153e-10
> sum(temp<-res1*kharif2$v46)
[1] -1.629371e-10
> sum(temp<-res1*kharif2$v45)
[1] -2.976677e-10
> sum(temp<-res1*log(kharif2$v34))
[1] 4.961032e-12
>
```

```
> sum(temp<-res2*rabi2$v37)
[1] -3.557218e-09
> sum(temp<-res2*log(rabi2$gdp))
[1] 1.177999e-10
> sum(temp<-res2*rabi2$tap)
[1] -1.359305e-09
> sum(temp<-res2*log(rabi2$beds))
[1] -2.033804e-09
> sum(temp<-res2*rabi2$v16)
[1] -3.419838e-09
> sum(temp<-res2*rabi2$v21)
[1] -1.927902e-10
> sum(temp<-res2*rabi2$v28)
[1] -3.285432e-10
> sum(temp<-res2*rabi2$v46)
[1] 8.151581e-11
> sum(temp<-res2*rabi2$v45)
[1] -3.716012e-10
> sum(temp<-res2*log(rabi2$v34))
[1] -1.273466e-10
>
```

One of the most important aspects of selecting a model is having an ideal R2 value, thus we minimize the ssr, which supports the fact that the sum of ei*xi is equal to 0 for each independent variable in our model.
As a result, orthogonality is also demonstrated.

Q2.

2a

```
measles<-EXmain$v36
yield_ind<-EXmain$index
model_base=lm(measles~yield_ind)
summary(model_base)
model_base$coefficients
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.419675   0.042466  80.528   <2e-16 ***
yield_ind   0.003823   0.002596   1.473    0.141
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.531 on 45968 degrees of freedom
  (27141 observations deleted due to missingness)
Multiple R-squared:  4.717e-05, Adjusted R-squared:  2.542e-0
5
F-statistic: 2.168 on 1 and 45968 DF,  p-value: 0.1409

> model_base$coefficients
(Intercept)   yield_ind
3.419675085 0.003823035
```

2b.

```
library(dplyr)
#true parameters
B_0 =3.419675  #intercept
B_1=0.003823    #slope
set.seed(1)
n=58489        #sample size
M=1000         #number of experiments

for (i in 1:M) {
  newdata<-EXmain %>% sample_frac(0.8)
  model_new=lm(newdata$v36~newdata$index)
  summary(model_new)
  model_new$coefficients
```

```
 ind_mean<-mean(newdata$index,na.rm = "True")
 ind_sd<-sd(newdata$index,na.rm = "True")
 err<-model_new$residuals
 err_mean<-mean(err,na.rm = "True")
 err_sd<-sd(err,na.rm = "True")

 U_i=rnorm(n,mean=err_mean,sd=err_sd)
 X_i=rnorm(n,mean = ind_mean,sd=ind_sd)
 Y_i=B_0+B_1*X_i+U_i
}
model_m1=lm(Y_i~X_i)
model_m1$coefficients
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.446989   0.037935  90.867   <2e-16 ***
X_i         0.003341   0.002314   1.444    0.149
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.576 on 58487 degrees of freedom
Multiple R-squared:  3.565e-05, Adjusted R-squared:  1.855e-0
5
F-statistic: 2.085 on 1 and 58487 DF,  p-value: 0.1487

> model_m1$coefficients
(Intercept)         X_i
3.446989009 0.003341081
```

Results for question 2 monte carlo simulations

In our question part (a) we had to run a simple linear regression where our independent variable was yield index and dependent variable was our given health indicator. Upon running this model we got the slopes and intercept values as
Slope:3.458922802
Intercept:0.003056363

As per the question we had to treat them as the true population parameters for the monte carlo simulations in 2b
After performing the simulations our values of slope and intercept came out to be close to the true values.
Slope:3.403909533
Intercept:0.00499046 {for instance 1}
Slope: 3.446989009 {for instance 2}
intercept:0.003341081
Which turned out to be close to the true values and hence it means that our estimates were consistent.

Q3.

(a)

```
main$nz <- c(0)
main$ez <- c(0)
main$wz <- c(0)
main$sz <- c(0)
main$cz <- c(0)
main$nez <- c(0)

for(i in 2:70572){
  s = main[i,4]
  if(s=="Himachal Pradesh" || s=="Punjab" || s=="Uttarakhand" || s=="Uttar Pradesh" || s=="Haryana") {
    main[i,98] = 1
  }
  if(s=="Bihar" || s=="Orissa" || s=="Jharkhand" || s=="West Bengal") {
    main[i,99] = 1
  }
  if(s=="Rajasthan" || s=="Gujarat" || s=="Goa" || s=="Maharashtra") {
    main[i,100] = 1
  }
  if(s=="Andhra Pradesh" || s=="Telangana" || s=="Karnataka" || s=="Kerala" || s=="Tamil Nadu") {
    main[i,101] = 1
  }
  if(s=="Madhya Pradesh" || s=="Chhattisgarh") {
    main[i,102] = 1
  }
  if(s=="Assam" || s=="Sikkim" || s=="Nagaland" || s=="Meghalaya" || s=="Manipur" || s=="Mizoram" || s=="Tripura" || s=="Arunachal Pradesh") {
    main[i,103] = 1
  }
}
```

```
model3 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 + lv34 + nz,data = kharif_data)
summary(model3)

model4 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 + lv34 + ez,data = kharif_data)
summary(model4)

model5 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 + lv34 + wz,data = kharif_data)
summary(model5)

model6 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 + lv34 + sz,data = kharif_data)
summary(model6)

model7 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 + lv34 + cz,data = kharif_data)
summary(model7)

model8 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 + lv34 + nez,data = kharif_data)
summary(model8)
```

(b)
In this question we were asked to use the t-statistic to test the null that there is no such structural break across different state-groups.

Our approach for this was to create 6 different models - each model contains our initial model along with one dummy variable for the different state groups.

KHARIF

```
> model3 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 +
 lv34 + nz,data = kharif_data)
> summary(model3)

Call:
lm(formula = v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 +
    v46 + v45 + lv34 + nz, data = kharif_data)

Residuals:
    Min      1Q  Median      3Q     Max
-23.057  -1.749  -0.424   1.337  56.418

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 15.995551   0.664930  24.056  < 2e-16 ***
lgdp        -1.159702   0.064811 -17.894  < 2e-16 ***
tap         -0.008760   0.001758  -4.983 6.31e-07 ***
lbeds        1.671311   0.071302  23.440  < 2e-16 ***
v37         -0.197049   0.002584 -76.272  < 2e-16 ***
v16          0.045793   0.003841  11.922  < 2e-16 ***
v21         -0.005062   0.001267  -3.996 6.47e-05 ***
v28         -0.056541   0.004424 -12.781  < 2e-16 ***
v46          0.577953   0.012269  47.106  < 2e-16 ***
v45          0.091541   0.004225  21.667  < 2e-16 ***
lv34         0.240418   0.052181   4.607 4.11e-06 ***
nz           2.229302   0.112084  19.890  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.128 on 17232 degrees of freedom
```

As can be seen above the p value of the North Zone state group is less than alpha=0.05 hence we can reject the null hypothesis of the test therefore $\widehat{\beta}\, North$ accounts for a structural break in the mean outcome level across designated state-groups.

```
> model4 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 +
  lv34 + ez,data = kharif_data)
> summary(model4)

Call:
lm(formula = v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 +
    v46 + v45 + lv34 + ez, data = kharif_data)

Residuals:
    Min      1Q  Median      3Q     Max
-21.844  -1.779  -0.393   1.256  57.958

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 17.419682   0.674446  25.828  < 2e-16 ***
lgdp        -1.297343   0.065687 -19.750  < 2e-16 ***
tap         -0.001737   0.001804  -0.963   0.3356
lbeds        2.024106   0.072149  28.055  < 2e-16 ***
v37         -0.197871   0.002616 -75.629  < 2e-16 ***
v16          0.031161   0.003826   8.145 4.04e-16 ***
v21         -0.007922   0.001273  -6.222 5.02e-10 ***
v28         -0.055230   0.004512 -12.240  < 2e-16 ***
v46          0.588044   0.012392  47.452  < 2e-16 ***
v45          0.102455   0.004236  24.184  < 2e-16 ***
lv34         0.117815   0.055809   2.111   0.0348 *
ez           0.622490   0.151551   4.107 4.02e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.184 on 17232 degrees of freedom
```

As can be seen above the p value of the North Zone state group is less than alpha=0.05 hence we can reject the null hypothesis of the test therefore $\widehat{\beta}\, East$ accounts for a structural break in the mean outcome level across designated state-groups.

```
> model5 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 +
 lv34 + wz,data = kharif_data)
> summary(model5)

Call:
lm(formula = v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 +
    v46 + v45 + lv34 + wz, data = kharif_data)

Residuals:
    Min      1Q  Median      3Q     Max
-21.897  -1.816  -0.362   1.299  57.868

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 17.135903   0.702058  24.408  < 2e-16 ***
lgdp        -1.270480   0.067043 -18.950  < 2e-16 ***
tap         -0.003541   0.001853  -1.911  0.05599 .
lbeds        1.963752   0.071170  27.593  < 2e-16 ***
v37         -0.197249   0.002614 -75.444  < 2e-16 ***
v16          0.028859   0.003821   7.552 4.48e-14 ***
v21         -0.007758   0.001307  -5.937 2.96e-09 ***
v28         -0.057940   0.004647 -12.469  < 2e-16 ***
v46          0.588754   0.012398  47.487  < 2e-16 ***
v45          0.102272   0.004239  24.129  < 2e-16 ***
lv34         0.193158   0.052727   3.663  0.00025 ***
wz           0.022434   0.120518   0.186  0.85233
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.187 on 17232 degrees of freedom
```

As can be seen above the p value of the North Zone state group is more than alpha=0.05 hence the null hypothesis of the test is not rejected therefore $\widehat{\beta}\, West$ does not account for a structural break in the mean outcome level across designated state-groups.

```
> model6 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 +
 lv34 + sz,data = kharif_data)
> summary(model6)

Call:
lm(formula = v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 +
    v46 + v45 + lv34 + sz, data = kharif_data)

Residuals:
    Min      1Q  Median      3Q     Max
-21.748  -1.807  -0.351   1.223  57.412

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.270635   0.673701  24.151  < 2e-16 ***
lgdp        -1.394024   0.066419 -20.988  < 2e-16 ***
tap         -0.002685   0.001754  -1.531 0.125896
lbeds        2.274064   0.077329  29.408  < 2e-16 ***
v37         -0.196568   0.002607 -75.409  < 2e-16 ***
v16          0.036989   0.003868   9.563  < 2e-16 ***
v21         -0.004516   0.001314  -3.436 0.000592 ***
v28         -0.070385   0.004647 -15.145  < 2e-16 ***
v46          0.587510   0.012364  47.518  < 2e-16 ***
v45          0.100378   0.004231  23.724  < 2e-16 ***
lv34         0.096622   0.053505   1.806 0.070961 .
sz          -1.220091   0.125196  -9.745  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.172 on 17232 degrees of freedom
```

As can be seen above the p value of the North Zone state group is less than alpha=0.05 hence we can reject the null hypothesis of the test therefore $\widehat{\beta}\,South$ accounts for a structural break in the mean outcome level across designated state-groups.

```
> model7 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 +
 lv34 + cz,data = kharif_data)
> summary(model7)

Call:
lm(formula = v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 +
    v46 + v45 + lv34 + cz, data = kharif_data)

Residuals:
    Min      1Q  Median      3Q     Max
-22.651  -1.803  -0.237   1.269  57.138

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 17.421849   0.662872  26.282  < 2e-16 ***
lgdp        -1.050606   0.065510 -16.037  < 2e-16 ***
tap         -0.008488   0.001756  -4.834 1.35e-06 ***
lbeds        1.692158   0.071108  23.797  < 2e-16 ***
v37         -0.203675   0.002604 -78.215  < 2e-16 ***
v16          0.025742   0.003750   6.864 6.94e-12 ***
v21         -0.008909   0.001260  -7.068 1.63e-12 ***
v28         -0.056246   0.004424 -12.713  < 2e-16 ***
v46          0.578898   0.012268  47.187  < 2e-16 ***
v45          0.093246   0.004215  22.121  < 2e-16 ***
lv34         0.217858   0.052146   4.178 2.96e-05 ***
cz          -2.442186   0.123202 -19.823  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.129 on 17232 degrees of freedom
```

As can be seen above the p value of the North Zone state group is less than alpha=0.05 hence we can reject the null hypothesis of the test therefore $\widehat{\beta}\ Central$ accounts for a structural break in the mean outcome level across designated state-groups.

```
> model8 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 +
 lv34 + nez,data = kharif_data)
> summary(model8)

Call:
lm(formula = v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 +
    v46 + v45 + lv34 + nez, data = kharif_data)

Residuals:
    Min      1Q  Median      3Q     Max
-22.944  -1.814  -0.323   1.335  56.787

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.370502   0.730831  14.190  < 2e-16 ***
lgdp        -1.298920   0.064468 -20.148  < 2e-16 ***
tap          0.007379   0.001804   4.089 4.35e-05 ***
lbeds        2.420805   0.072807  33.249  < 2e-16 ***
v37         -0.207049   0.002618 -79.088  < 2e-16 ***
v16          0.027413   0.003740   7.330 2.40e-13 ***
v21         -0.008925   0.001258  -7.096 1.33e-12 ***
v28         -0.058430   0.004415 -13.236  < 2e-16 ***
v46          0.578026   0.012242  47.215  < 2e-16 ***
v45          0.089096   0.004226  21.083  < 2e-16 ***
lv34         0.459497   0.053459   8.595  < 2e-16 ***
nez          3.400158   0.157293  21.617  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.118 on 17232 degrees of freedom
```

As can be seen above the p value of the North Zone state group is less than alpha=0.05 hence we can reject the null hypothesis of the test therefore $\widehat{\beta}\, NorthEast$ accounts for a structural break in the mean outcome level across designated state-groups.

RABI

```
> model3 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 +
  lv34 + nz,data = rabi_data)
> summary(model3)

Call:
lm(formula = v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 +
    v46 + v45 + lv34 + nz, data = rabi_data)

Residuals:
    Min      1Q  Median      3Q     Max
-21.779  -1.868  -0.423   1.302  57.172

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 13.743758   0.700517  19.619  < 2e-16 ***
lgdp        -1.208295   0.068574 -17.620  < 2e-16 ***
tap         -0.008443   0.001909  -4.424 9.77e-06 ***
lbeds        1.750552   0.075022  23.334  < 2e-16 ***
v37         -0.184102   0.002621 -70.244  < 2e-16 ***
v16          0.044917   0.003979  11.289  < 2e-16 ***
v21         -0.004946   0.001350  -3.663  0.00025 ***
v28         -0.065643   0.004973 -13.201  < 2e-16 ***
v46          0.584595   0.013630  42.890  < 2e-16 ***
v45          0.095303   0.004520  21.083  < 2e-16 ***
lv34         0.355526   0.057317   6.203 5.69e-10 ***
nz           2.214273   0.119790  18.485  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.232 on 15392 degrees of freedom
```

As can be seen above the p value of the North Zone state group is less than alpha=0.05 hence we can reject the null hypothesis of the test therefore $\widehat{\beta}\, North$ accounts for a structural break in the mean outcome level across designated state-groups.

```
> model4 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 +
  lv34 + ez,data = rabi_data)
> summary(model4)

Call:
lm(formula = v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 +
    v46 + v45 + lv34 + ez, data = rabi_data)

Residuals:
    Min      1Q  Median      3Q     Max
-20.727  -1.897  -0.382   1.248  58.795

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.773832   0.711526  20.764  < 2e-16 ***
lgdp        -1.336853   0.069560 -19.219  < 2e-16 ***
tap         -0.001706   0.001971  -0.866 0.386740
lbeds        2.116987   0.075945  27.875  < 2e-16 ***
v37         -0.183563   0.002649 -69.302  < 2e-16 ***
v16          0.032253   0.003972   8.119 5.05e-16 ***
v21         -0.007777   0.001356  -5.737 9.84e-09 ***
v28         -0.063600   0.005090 -12.494  < 2e-16 ***
v46          0.595712   0.013761  43.289  < 2e-16 ***
v45          0.106695   0.004528  23.562  < 2e-16 ***
lv34         0.211818   0.061529   3.443 0.000578 ***
ez           0.503184   0.142795   3.524 0.000427 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.288 on 15392 degrees of freedom
```

As can be seen above the p value of the North Zone state group is less than alpha=0.05 hence we can reject the null hypothesis of the test therefore $\widehat{\beta}\, East$ accounts for a structural break in the mean outcome level across designated state-groups.

```
> model5 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 +
 lv34 + wz,data = rabi_data)
> summary(model5)

Call:
lm(formula = v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 +
    v46 + v45 + lv34 + wz, data = rabi_data)

Residuals:
    Min      1Q  Median      3Q     Max
-20.841  -1.939  -0.379   1.281  58.667

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.471497   0.737964  19.610  < 2e-16 ***
lgdp        -1.307289   0.070906 -18.437  < 2e-16 ***
tap         -0.003381   0.002036  -1.661   0.0967 .
lbeds        2.055520   0.074573  27.564  < 2e-16 ***
v37         -0.183433   0.002653 -69.152  < 2e-16 ***
v16          0.030608   0.003961   7.728 1.16e-14 ***
v21         -0.007881   0.001388  -5.678 1.39e-08 ***
v28         -0.066373   0.005193 -12.782  < 2e-16 ***
v46          0.595927   0.013767  43.285  < 2e-16 ***
v45          0.106444   0.004530  23.499  < 2e-16 ***
lv34         0.286183   0.057830   4.949 7.55e-07 ***
wz          -0.009079   0.138029  -0.066   0.9476
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.29 on 15392 degrees of freedom
```

As can be seen above the p value of the North Zone state group is more than alpha=0.05 hence the null hypothesis of the test is not rejected therefore $\widehat{\beta}\, West$ does not account for a structural break in the mean outcome level across designated state-groups.

```
> model6 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 +
 lv34 + sz,data = rabi_data)
> summary(model6)

Call:
lm(formula = v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 +
    v46 + v45 + lv34 + sz, data = rabi_data)

Residuals:
    Min      1Q  Median      3Q     Max
-20.515  -1.923  -0.377   1.215  58.150

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 13.678822   0.710286  19.258  < 2e-16 ***
lgdp        -1.435662   0.070240 -20.439  < 2e-16 ***
tap         -0.002192   0.001909  -1.148 0.250916
lbeds        2.380239   0.081452  29.223  < 2e-16 ***
v37         -0.183206   0.002642 -69.340  < 2e-16 ***
v16          0.038663   0.004027   9.600  < 2e-16 ***
v21         -0.004679   0.001394  -3.357 0.000791 ***
v28         -0.079635   0.005206 -15.297  < 2e-16 ***
v46          0.594246   0.013729  43.285  < 2e-16 ***
v45          0.104316   0.004522  23.067  < 2e-16 ***
lv34         0.181957   0.058718   3.099 0.001947 **
sz          -1.273147   0.135608  -9.388  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.275 on 15392 degrees of freedom
```

As can be seen above the p value of the North Zone state group is less than alpha=0.05 hence we can reject the null hypothesis of the test therefore $\widehat{\beta}\,South$ accounts for a structural break in the mean outcome level across designated state-groups.

```
> model7 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 +
 lv34 + cz,data = rabi_data)
> summary(model7)

Call:
lm(formula = v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 +
    v46 + v45 + lv34 + cz, data = rabi_data)

Residuals:
    Min      1Q  Median      3Q     Max
-21.205  -1.933  -0.251   1.252  58.304

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.670712   0.702801  20.875  < 2e-16 ***
lgdp        -1.165636   0.069444 -16.785  < 2e-16 ***
tap         -0.006752   0.001913  -3.529 0.000418 ***
lbeds        1.875564   0.074659  25.122  < 2e-16 ***
v37         -0.186283   0.002641 -70.537  < 2e-16 ***
v16          0.029666   0.003922   7.565 4.11e-14 ***
v21         -0.009081   0.001350  -6.725 1.82e-11 ***
v28         -0.064727   0.004998 -12.952  < 2e-16 ***
v46          0.590475   0.013686  43.143  < 2e-16 ***
v45          0.100402   0.004522  22.203  < 2e-16 ***
lv34         0.288211   0.057465   5.015 5.35e-07 ***
cz          -1.922851   0.138085 -13.925  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.257 on 15392 degrees of freedom
```

As can be seen above the p value of the North Zone state group is less than alpha=0.05 hence we can reject the null hypothesis of the test therefore $\widehat{\beta}\, Central$ accounts for a structural break in the mean outcome level across designated state-groups.

```
> model8 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 +
 lv34 + nez,data = rabi_data)
> summary(model8)

Call:
lm(formula = v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 +
    v46 + v45 + lv34 + nez, data = rabi_data)

Residuals:
    Min      1Q  Median      3Q     Max
-21.884  -1.981  -0.362   1.324  57.685

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.446162   0.773231  10.923  < 2e-16 ***
lgdp        -1.314306   0.068374 -19.222  < 2e-16 ***
tap          0.004762   0.001942   2.452   0.0142 *
lbeds        2.389567   0.075419  31.684  < 2e-16 ***
v37         -0.192847   0.002671 -72.203  < 2e-16 ***
v16          0.032547   0.003905   8.335  < 2e-16 ***
v21         -0.008344   0.001342  -6.219 5.14e-10 ***
v28         -0.066743   0.004973 -13.420  < 2e-16 ***
v46          0.585935   0.013630  42.989  < 2e-16 ***
v45          0.093853   0.004533  20.703  < 2e-16 ***
lv34         0.553622   0.059037   9.378  < 2e-16 ***
nez          3.205922   0.174932  18.327  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.233 on 15392 degrees of freedom
```

As can be seen above the p value of the North Zone state group is less than alpha=0.05 hence we can reject the null hypothesis of the test therefore $\widehat{\beta}\, NorthEast$ accounts for a structural break in the mean outcome level across designated state-groups.

(c)

KHARIF

```
model12 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 + lv34 + sz,data = kharif2)
summary(model12)
library(car)
nullhyp <- c("sz")
linearHypothesis(model12,nullhyp)
```

```
> library(car)
> nullhyp <- c("sz")
> linearHypothesis(model12,nullhyp)
Linear hypothesis test

Hypothesis:
sz = 0

Model 1: restricted model
Model 2: v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 +
    lv34 + sz

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1  17233 463569
2  17232 461028  1    2540.9 94.974 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

According to the ANOVA table, the value corresponding to the Df and residual Df values given here is 2.7055 which is less than the F value (=94.974) here. Thus, we can reject the null hypothesis that there is no structural break across southern and non-southern state-groups. The same result was also obtained from the t-statistic test for kharif data and hence, can be verified.

RABI

```
model11 <- lm(v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 + lv34 + sz,data = rabi2)
summary(model11)
library(car)
nullhyp <- c("sz")
linearHypothesis(model11,nullhyp)
```

```
> library(car)
> nullhyp <- c("sz")
> linearHypothesis(model11,nullhyp)
Linear hypothesis test

Hypothesis:
sz = 0

Model 1: restricted model
Model 2: v36 ~ lgdp + tap + lbeds + v37 + v16 + v21 + v28 + v46 + v45 +
    lv34 + sz

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1  15393 430765
2  15392 428312  1    2452.8 88.143 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

According to the ANOVA table, the value corresponding to the Df and residual Df values given here is 2.7055 which is less than the F value (=88.143) here. Thus, we can reject the null hypothesis that there is no structural break across southern and non-southern state-groups. The same result was also obtained from the t-statistic test for rabi data and hence, can be verified.