601 group3 project2 models

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
rm(list = ls())
data <- read.csv('modefieddata.csv')
names(data)
```

```
## [1] "state"             "quarter"           "new.cases"
## [4] "death"             "lockdown.days"     "GDP.USD."
## [7] "Personal.Income"   "unemployment.rate" "vaccination.rate"
```

```
data$state <- as.factor(data$state)
data$quarter <- as.factor(data$quarter)
data$unemployment.rate <- as.numeric(data$unemployment.rate)
data$new.cases <- as.numeric(data$new.cases)
data$death <- as.numeric(data$death)
data$lockdown.days <- as.numeric(data$lockdown.days)
data$vaccination.rate <- as.numeric(data$vaccination.rate)
data$GDP.USD. <- as.numeric(data$GDP.USD.)
data$Personal.Income <- as.numeric(data$Personal.Income)
```

```
library(lmtest)
```

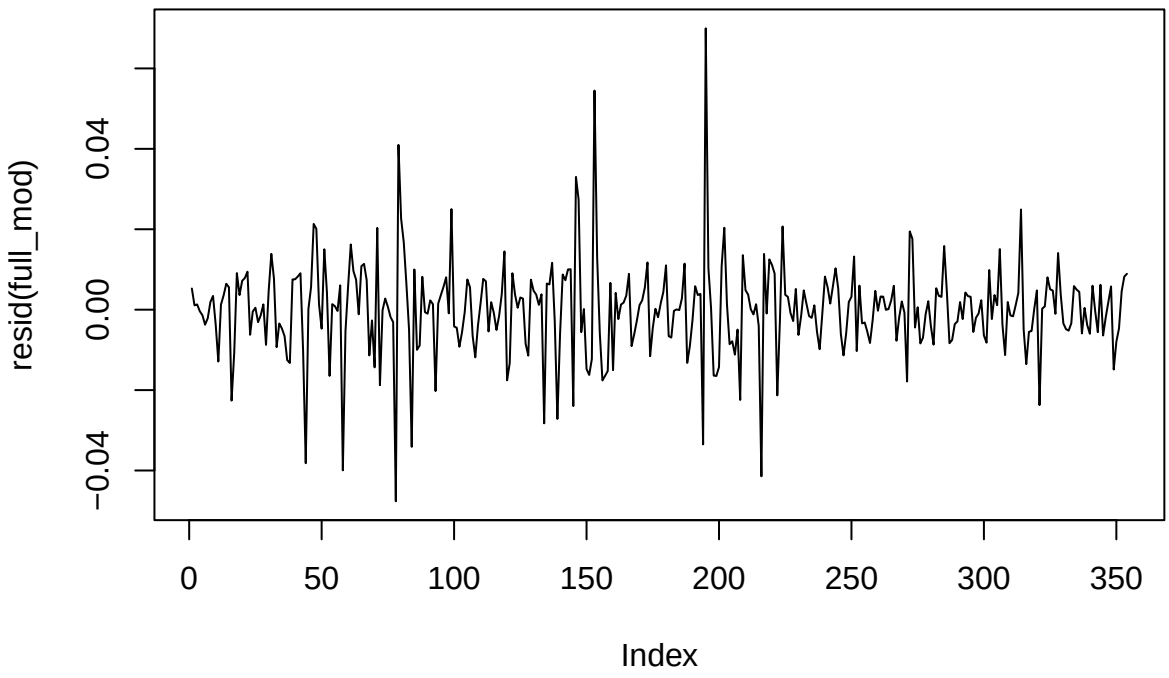
```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
full_mod<- lm(unemployment.rate ~., data = data)
summary(full_mod)
```

```
##
## Call:
## lm(formula = unemployment.rate ~ ., data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.047643 -0.005401  0.000459  0.005547  0.069967
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 6.791e-03  8.406e-03   0.808 0.419829
## stateAlaska                 3.844e-02  8.955e-03   4.292 2.41e-05 ***
## stateArizona                1.202e-02  8.135e-03   1.478 0.140550
## stateArkansas               1.379e-02  7.326e-03   1.882 0.060794 .
## stateCalifornia            -1.482e-01  8.254e-02  -1.796 0.073566 .
## stateColorado               5.888e-03  8.360e-03   0.704 0.481837
## stateConnecticut            1.744e-02  7.444e-03   2.343 0.019782 *
## stateDelaware               3.449e-02  8.805e-03   3.917 0.000112 ***
## stateDistrict of Columbia   3.701e-02  8.674e-03   4.267 2.68e-05 ***
## stateFlorida               -6.017e-02  3.233e-02  -1.861 0.063764 .
## stateGeorgia               -2.006e-02  1.253e-02  -1.601 0.110359
## stateHawaii                 5.907e-02  8.299e-03   7.118 8.59e-12 ***
## stateIdaho                  3.685e-03  8.092e-03   0.455 0.649179
## stateIllinois              -1.558e-02  1.975e-02  -0.789 0.430826
## stateIndiana               -1.874e-03  7.918e-03  -0.237 0.813035
## stateIowa                   3.627e-03  7.101e-03   0.511 0.609910
## stateKansas                 4.083e-03  7.121e-03   0.573 0.566833
## stateKentucky               8.062e-03  6.882e-03   1.171 0.242358
## stateLouisiana              2.866e-02  6.811e-03   4.208 3.44e-05 ***
## stateMaine                  1.534e-02  9.167e-03   1.673 0.095393 .
## stateMaryland               4.263e-04  9.059e-03   0.047 0.962498
## stateMassachusetts          7.930e-03  1.231e-02   0.644 0.519834
## stateMichigan               6.278e-03  1.191e-02   0.527 0.598496
## stateMinnesota             -8.368e-03  7.986e-03  -1.048 0.295630
## stateMississippi            2.902e-02  7.552e-03   3.842 0.000150 ***
## stateMissouri              -4.415e-03  7.412e-03  -0.596 0.551905
## stateMontana                1.161e-02  8.713e-03   1.333 0.183628
## stateNebraska              -6.781e-03  7.784e-03  -0.871 0.384414
## stateNevada                 6.165e-02  7.073e-03   8.716 2.23e-16 ***
## stateNew Hampshire          1.107e-02  8.202e-03   1.350 0.178186
## stateNew Jersey             4.746e-03  1.537e-02   0.309 0.757776
## stateNew Mexico             4.107e-02  8.101e-03   5.070 7.09e-07 ***
## stateNew York              -4.919e-02  3.988e-02  -1.233 0.218449
## stateNorth Carolina        -1.164e-02  1.206e-02  -0.965 0.335191
## stateNorth Dakota           1.056e-02  8.925e-03   1.183 0.237606
## stateOhio                  -1.217e-02  1.461e-02  -0.833 0.405607
## stateOklahoma               7.442e-03  6.879e-03   1.082 0.280227
## stateOregon                 1.584e-02  7.000e-03   2.262 0.024416 *
## statePennsylvania          -9.848e-03  1.920e-02  -0.513 0.608397
## stateRhode Island           4.374e-02  8.747e-03   5.000 9.92e-07 ***
```

```
## stateSouth Carolina        3.914e-03  6.845e-03   0.572 0.567937
## stateSouth Dakota          6.363e-03  8.843e-03   0.720 0.472353
## stateTennessee             4.104e-03  7.886e-03   0.520 0.603114
## stateTexas                -8.243e-02  4.545e-02  -1.814 0.070762 .
## stateUtah                 -4.587e-03  7.060e-03  -0.650 0.516394
## stateVermont               1.063e-02  9.321e-03   1.140 0.255187
## stateVirginia             -2.046e-02  1.207e-02  -1.695 0.091212 .
## stateWashington           -2.108e-03  1.182e-02  -0.178 0.858571
## stateWest Virginia         3.231e-02  8.269e-03   3.908 0.000116 ***
## stateWisconsin            -3.994e-03  7.569e-03  -0.528 0.598073
## stateWyoming               2.216e-02  9.154e-03   2.421 0.016094 *
## quarter20-q2               6.998e-02  4.015e-03  17.427  < 2e-16 ***
## quarter20-q3               4.250e-02  2.610e-03  16.281  < 2e-16 ***
## quarter20-q4               2.454e-02  2.997e-03   8.189 8.40e-15 ***
## quarter21-q1               1.608e-02  4.452e-03   3.613 0.000357 ***
## quarter21-q2               1.442e-02  8.808e-03   1.637 0.102810
## quarter21-q3               1.024e-02  9.700e-03   1.056 0.291994
## new.cases                 -6.493e-09  1.019e-08  -0.637 0.524390
## death                      3.247e-07  4.993e-07   0.650 0.516058
## lockdown.days              3.149e-04  8.902e-05   3.537 0.000470 ***
## GDP.USD.                   2.712e-09  2.383e-09   1.138 0.256040
## Personal.Income            6.914e-08  3.191e-08   2.167 0.031070 *
## vaccination.rate          -7.678e-04  1.827e-02  -0.042 0.966508
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01271 on 291 degrees of freedom
## Multiple R-squared:  0.8728, Adjusted R-squared:  0.8457
## F-statistic: 32.21 on 62 and 291 DF,  p-value: < 2.2e-16
```

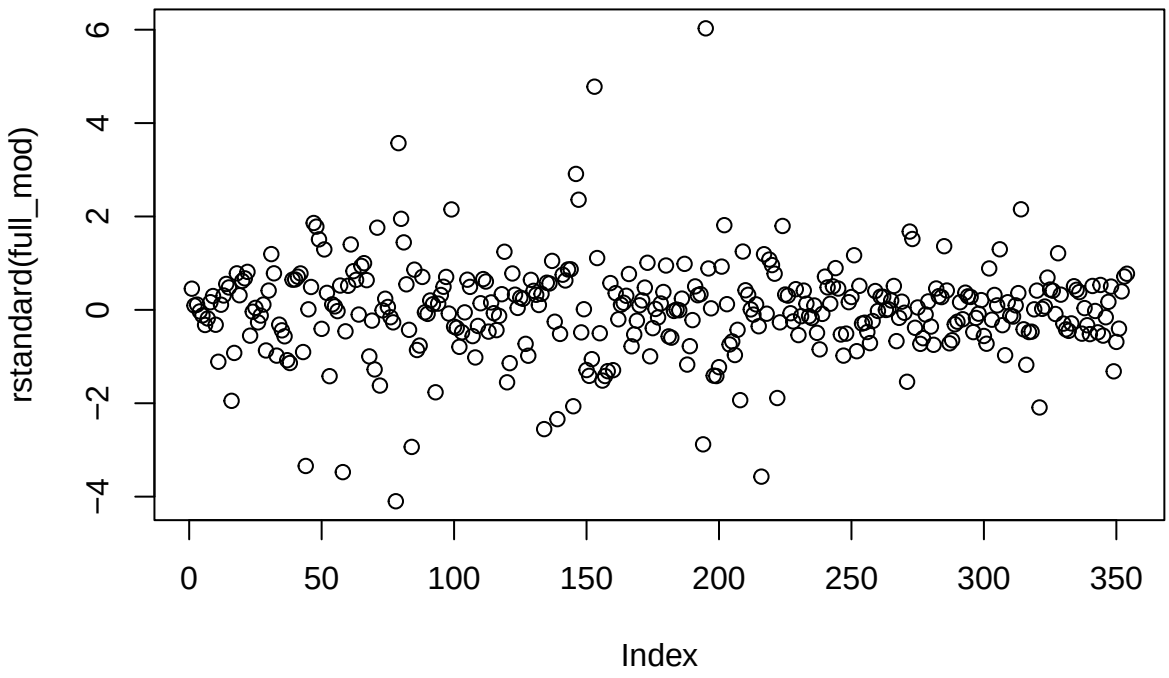
```
plot(resid(full_mod),type = "l")
```



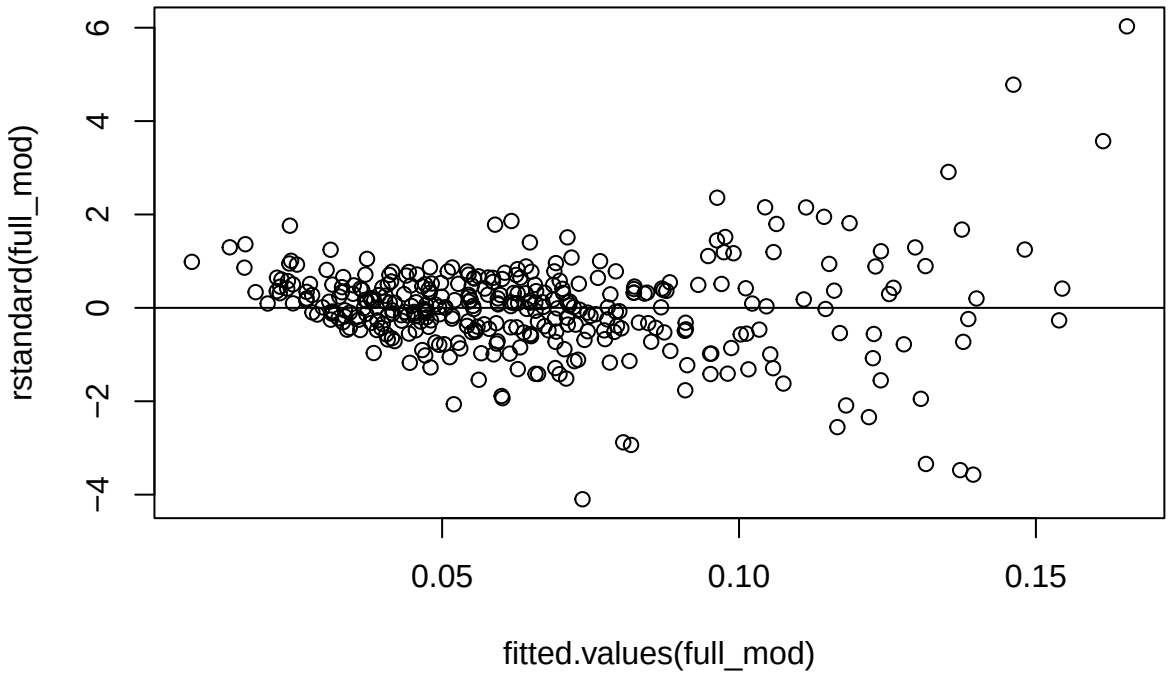

```
dwtest(full_mod)
```

```
##
##  Durbin-Watson test
##
## data:  full_mod
## DW = 1.9676, p-value = 0.0009053
## alternative hypothesis: true autocorrelation is greater than 0
```

```
plot(rstandard(full_mod), type = "p")#heteroscedasticity validation
```

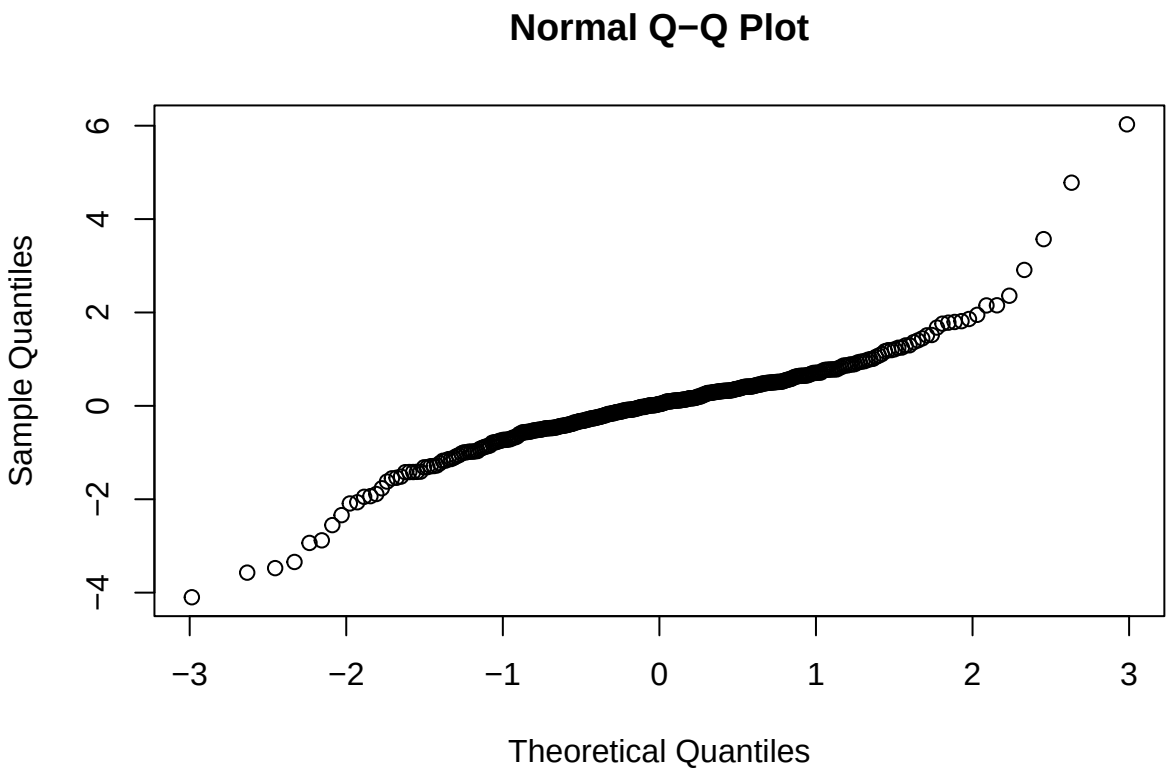



```
plot(fitted.values(full_mod), rstandard(full_mod)) +abline(h = 0) # linearity validation
```




```
## integer(0)
```

```
qqnorm(rstandard(full_mod))# normality validation
```




```
data$state <- as.character(data$state)
data$state[data$state == "Alabama"] <-"Southeast"
data$state[data$state == "Alaska"] <-"West"
data$state[data$state == "Arizona"] <-"Southwest"
data$state[data$state == "Arkansas"] <-"Southeast"
data$state[data$state == "California"] <-"West"
data$state[data$state == "Alabama"] <-"Southeast"
data$state[data$state == "Alaska"] <-"West"
data$state[data$state == "Arizona"] <-"Southwest"
data$state[data$state == "Arkansas"] <-"Southeast"
data$state[data$state == "California"] <-"West"
data$state[data$state == "Colorado"] <-"West"
data$state[data$state == "Connecticut"] <-"Northeast"
data$state[data$state == "District of Columbia"] <-"Southeast"
data$state[data$state == "Delaware"] <-"Midwest"
data$state[data$state == "Florida"] <-"Southeast"
data$state[data$state == "Georgia"] <-"Southeast"
data$state[data$state == "Hawaii"] <-"West"
data$state[data$state == "Idaho"] <-"West"
data$state[data$state == "Illinois"] <-"Midwest"
data$state[data$state == "Indiana"] <-"Midwest"
data$state[data$state == "Iowa"] <-"Midwest"
data$state[data$state == "Kansas"] <-"Midwest"
data$state[data$state == "Kentucky"] <-"Southeast"
data$state[data$state == "Louisiana"] <-"Southeast"
data$state[data$state == "Maine"] <-"Northeast"
```

```r
data$state[data$state == "Maryland"] <-"Northeast"
data$state[data$state == "Massachusetts"] <-"Northeast"
data$state[data$state == "Michigan"] <-"Midwest"
data$state[data$state == "Minnesota"] <-"Midwest"
data$state[data$state == "Mississippi"] <-"Midwest"
data$state[data$state == "Missouri"] <-"Midwest"
data$state[data$state == "Montana"] <-"West"
data$state[data$state == "Nebraska"] <-"Midwest"
data$state[data$state == "Nevada"] <-"West"
data$state[data$state == "New Hampshire"] <-"Northeast"
data$state[data$state == "New Jersey"] <-"Northeast"
data$state[data$state == "New Mexico"] <-"Southwest"
data$state[data$state == "New York"] <-"Northeast"
data$state[data$state == "North Carolina"] <-"Southeast"
data$state[data$state == "North Dakota"] <-"Midwest"
data$state[data$state == "Ohio"] <-"Midwest"
data$state[data$state == "Oklahoma"] <-"Southwest"
data$state[data$state == "Oregon"] <-"West"
data$state[data$state == "Pennsylvania"] <-"Northeast"
data$state[data$state == "Rhode Island"] <-"Northeast"
data$state[data$state == "South Carolina"] <-"Southeast"
data$state[data$state == "South Dakota"] <-"Midwest"
data$state[data$state == "Tennessee"] <-"Southeast"
data$state[data$state == "Texas"] <-"Southwest"
data$state[data$state == "Utah"] <-"West"
data$state[data$state == "Vermont"] <-"Northeast"
data$state[data$state == "Virginia"] <-"Southeast"
data$state[data$state == "Washington"] <-"West"
data$state[data$state == "West Virginia"] <-"Southeast"
data$state[data$state == "Wisconsin"] <-"Midwest"
data$state[data$state == "Wyoming"] <-"West"
fit1 = lm(unemployment.rate ~ ., data = data)
summary(fit1)
```

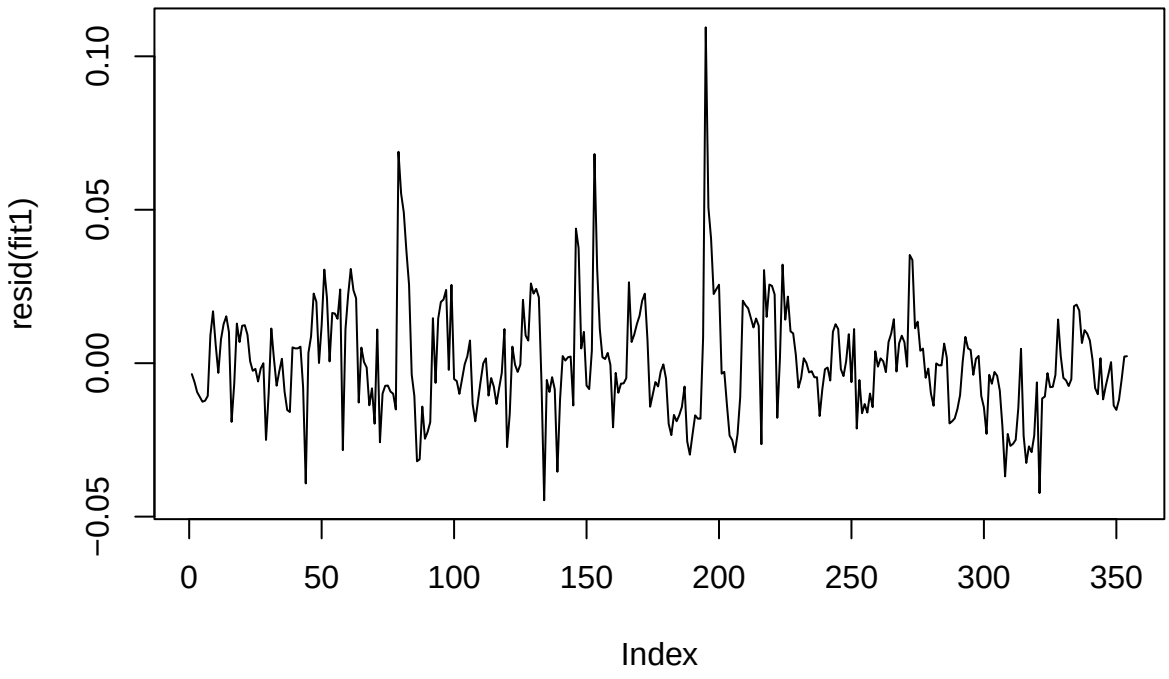
```
##
## Call:
## lm(formula = unemployment.rate ~ ., data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.044662 -0.010356 -0.001959  0.009637  0.109437
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.583e-02  3.133e-03   8.243 3.78e-15 ***
## stateNortheast    1.093e-02  3.075e-03   3.554 0.000434 ***
## stateSoutheast    3.642e-03  2.749e-03   1.325 0.186104
## stateSouthwest    9.843e-03  3.969e-03   2.480 0.013620 *
## stateWest         1.079e-02  2.809e-03   3.840 0.000147 ***
## quarter20-q2      6.358e-02  5.342e-03  11.902  < 2e-16 ***
## quarter20-q3      4.432e-02  3.692e-03  12.006  < 2e-16 ***
## quarter20-q4      2.642e-02  4.232e-03   6.243 1.29e-09 ***
## quarter21-q1      1.736e-02  5.775e-03   3.005 0.002852 **
```

```
## quarter21-q2      8.815e-03  1.142e-02   0.772 0.440905
## quarter21-q3      4.330e-03  1.256e-02   0.345 0.730451
## new.cases        -1.283e-08  1.388e-08  -0.925 0.355807
## death             8.349e-07  6.885e-07   1.213 0.226115
## lockdown.days     5.316e-04  1.167e-04   4.557 7.28e-06 ***
## GDP.USD.          2.883e-09  3.148e-09   0.916 0.360445
## Personal.Income   6.353e-09  4.612e-09   1.378 0.169266
## vaccination.rate  1.791e-02  2.342e-02   0.765 0.445051
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0182 on 337 degrees of freedom
## Multiple R-squared:  0.6979, Adjusted R-squared:  0.6836
## F-statistic: 48.66 on 16 and 337 DF,  p-value: < 2.2e-16
```

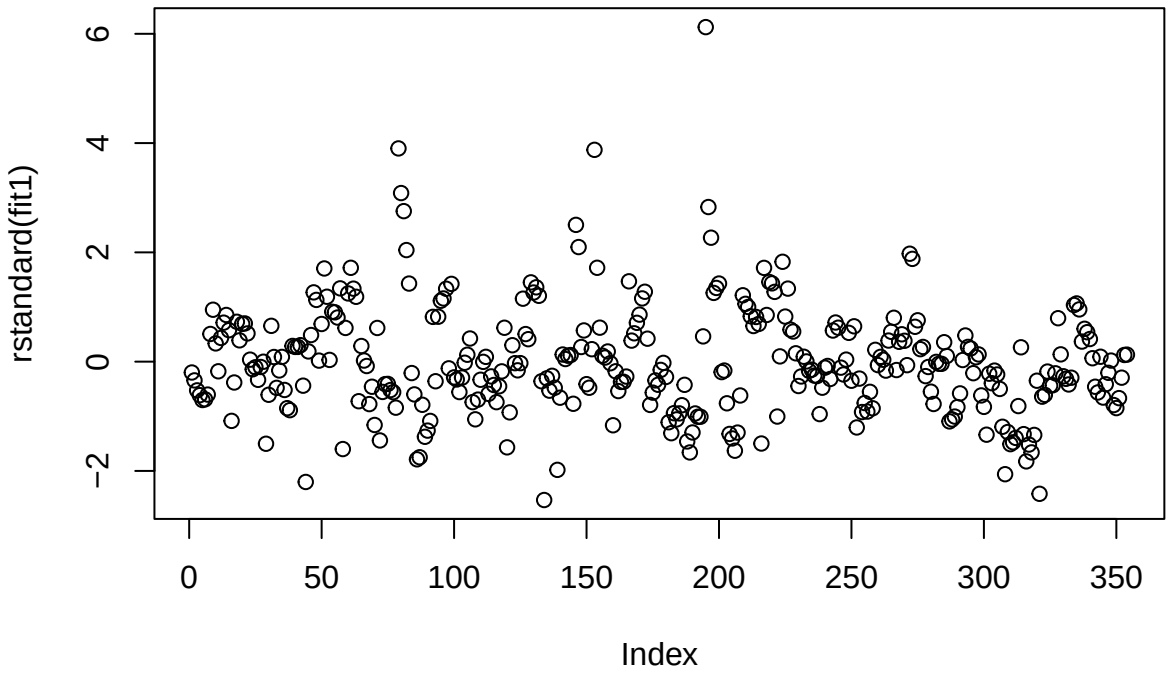
```
dwtest(fit1)
```

```
##
##  Durbin-Watson test
##
## data:  fit1
## DW = 0.86771, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

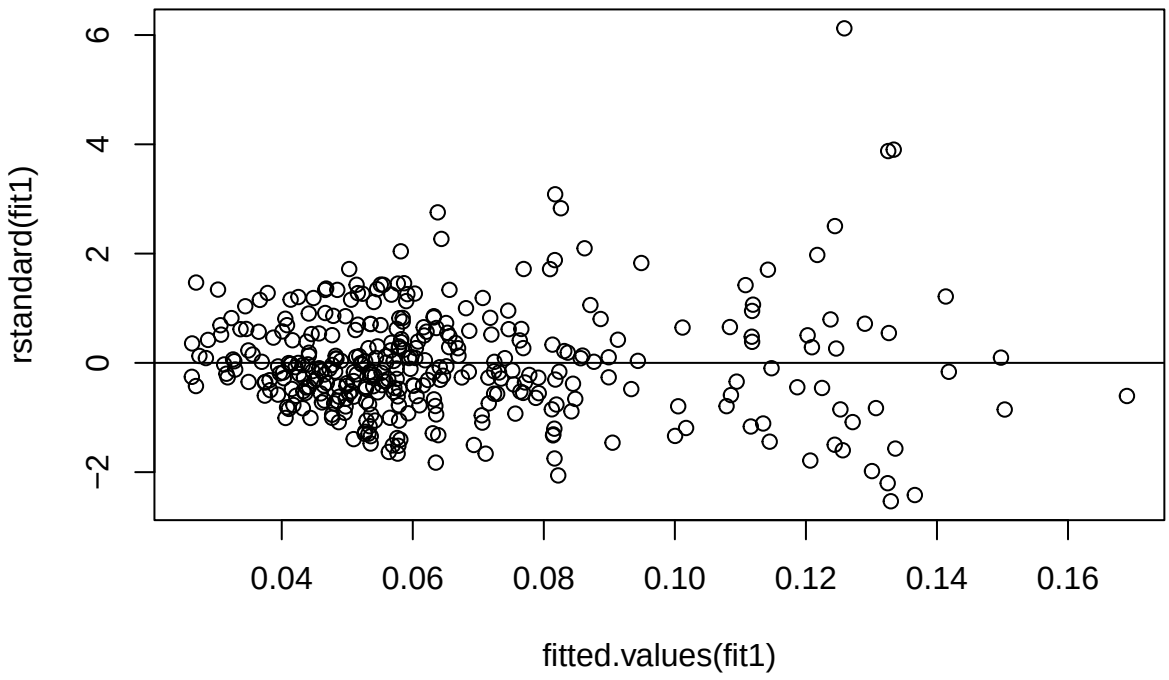
```
plot(resid(fit1),type = "l")
```



```r
plot(rstandard(fit1), type = "p")#heteroscedasticity validation
```

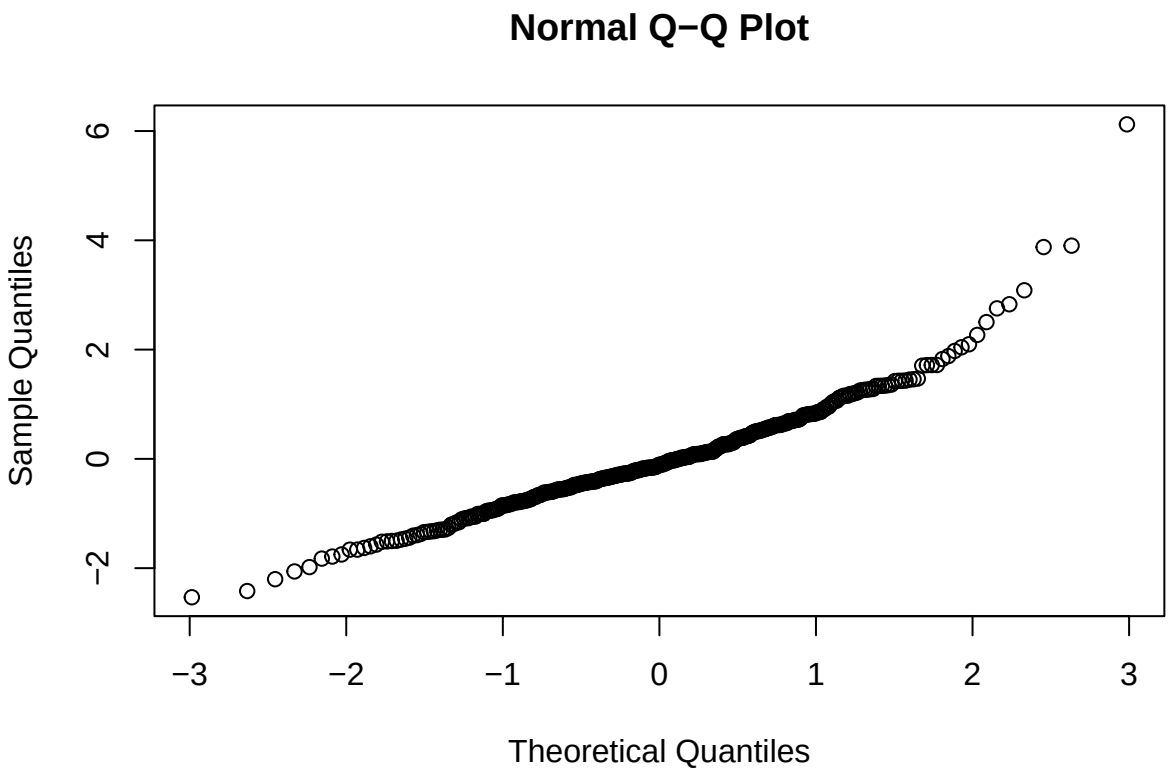



```r
plot(fitted.values(fit1), rstandard(fit1)) +abline(h = 0) # linearity validation
```

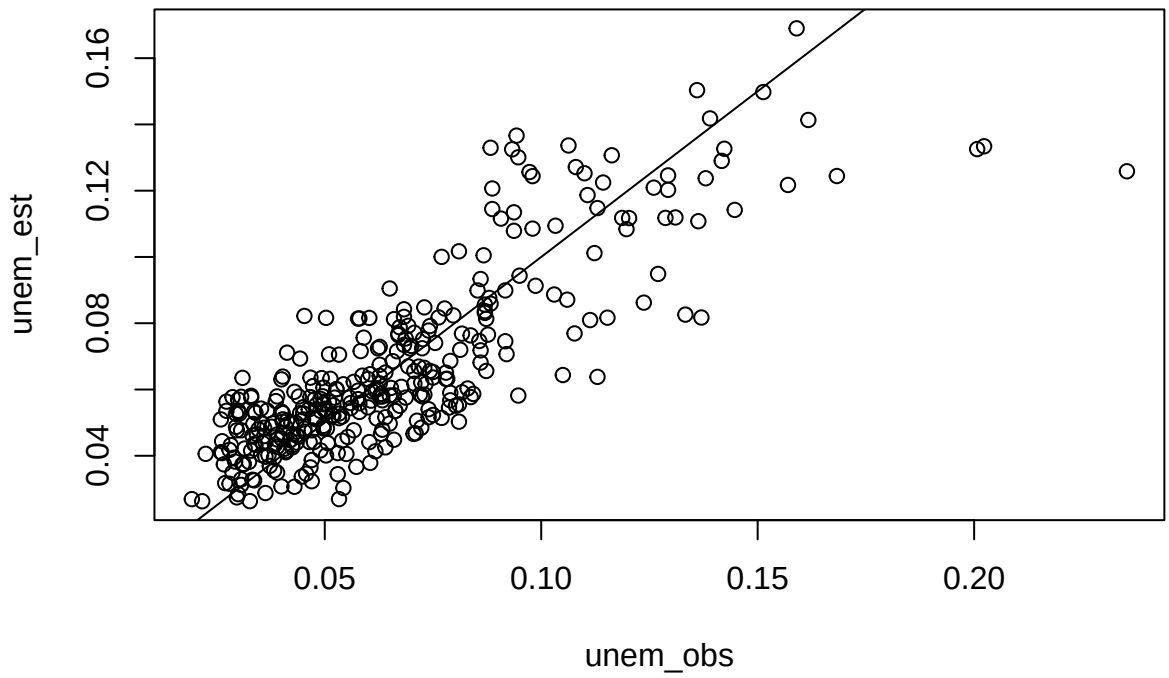



```
## integer(0)
```

```
qqnorm(rstandard(fit1))# normality validation
```




```
unem_obs<-data[,8]
unem_est<-predict(fit1,type="response")
plot(unem_obs,unem_est)+abline(a=0,b=1)
```




```
## integer(0)
```

```
logit_mod<- glm(unemployment.rate ~., family=binomial(link="logit"),data = data)
```

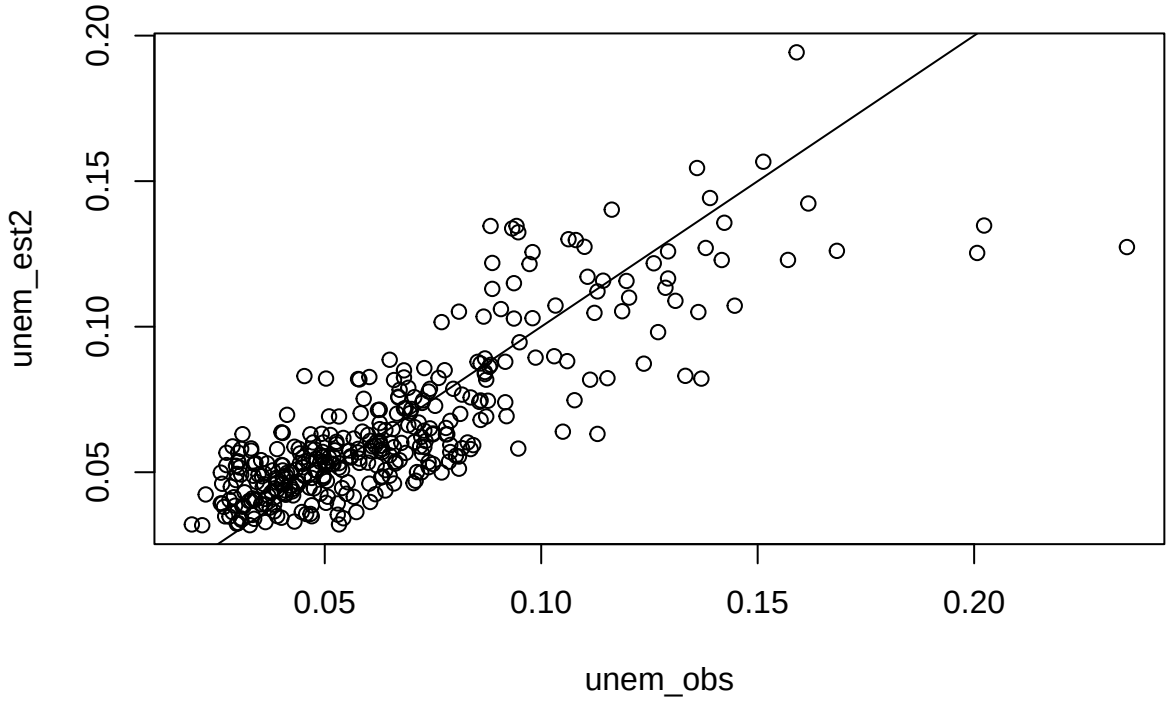
```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

```
summary(logit_mod)
```

```
##
## Call:
## glm(formula = unemployment.rate ~ ., family = binomial(link = "logit"),
##     data = data)
##
## Deviance Residuals:
##       Min         1Q     Median         3Q        Max
## -0.148654  -0.045862  -0.008329   0.042160   0.294856
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -3.423e+00  8.518e-01  -4.018 5.87e-05 ***
## stateNortheast    1.867e-01  6.844e-01   0.273    0.785
## stateSoutheast    6.960e-02  6.386e-01   0.109    0.913
## stateSouthwest    1.748e-01  8.772e-01   0.199    0.842
## stateWest         1.821e-01  6.388e-01   0.285    0.776
```

```
## quarter20-q2      1.077e+00  1.126e+00   0.956    0.339
## quarter20-q3      8.163e-01  9.150e-01   0.892    0.372
## quarter20-q4      5.330e-01  1.064e+00   0.501    0.616
## quarter21-q1      3.539e-01  1.454e+00   0.243    0.808
## quarter21-q2      1.473e-01  2.854e+00   0.052    0.959
## quarter21-q3      4.127e-02  3.131e+00   0.013    0.989
## new.cases        -3.944e-08  2.744e-06  -0.014    0.989
## death             3.993e-06  1.256e-04   0.032    0.975
## lockdown.days     4.812e-03  2.024e-02   0.238    0.812
## GDP.USD.          5.350e-08  6.672e-07   0.080    0.936
## Personal.Income   7.435e-08  9.478e-07   0.078    0.937
## vaccination.rate  4.735e-01  5.735e+00   0.083    0.934
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 5.4174  on 353  degrees of freedom
## Residual deviance: 1.6039  on 337  degrees of freedom
## AIC: 82.126
##
## Number of Fisher Scoring iterations: 6
```

```
#library(arm)
#binnedplot(fitted(logit_mod),resid(logit_mod,type="response"))
#the observed data against the estimated data
unem_obs<-data[,8]
unem_est2<-predict(logit_mod,type="response")
plot(unem_obs,unem_est2)+abline(a=0,b=1)
```




```
## integer(0)
```

```
#check the dispersion of the model
quasi_lm<-glm(unemployment.rate ~., family=quasibinomial(link="logit"),data = data)
summary(quasi_lm)
```

```
##
## Call:
## glm(formula = unemployment.rate ~ ., family = quasibinomial(link = "logit"),
##     data = data)
##
## Deviance Residuals:
##       Min         1Q     Median         3Q        Max
## -0.148654  -0.045862  -0.008329   0.042160   0.294856
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -3.423e+00  5.932e-02 -57.702  < 2e-16 ***
## stateNortheast    1.867e-01  4.766e-02   3.917 0.000109 ***
## stateSoutheast    6.960e-02  4.447e-02   1.565 0.118484
## stateSouthwest    1.748e-01  6.108e-02   2.862 0.004472 **
## stateWest         1.821e-01  4.448e-02   4.093 5.33e-05 ***
## quarter20-q2      1.077e+00  7.840e-02  13.735  < 2e-16 ***
## quarter20-q3      8.163e-01  6.372e-02  12.812  < 2e-16 ***
## quarter20-q4      5.330e-01  7.407e-02   7.197 4.03e-12 ***
```

```
## quarter21-q1      3.539e-01  1.013e-01   3.495 0.000537 ***
## quarter21-q2      1.473e-01  1.988e-01   0.741 0.459020
## quarter21-q3      4.127e-02  2.180e-01   0.189 0.850003
## new.cases        -3.944e-08  1.911e-07  -0.206 0.836613
## death             3.993e-06  8.749e-06   0.456 0.648419
## lockdown.days     4.812e-03  1.410e-03   3.414 0.000719 ***
## GDP.USD.          5.350e-08  4.646e-08   1.151 0.250382
## Personal.Income   7.435e-08  6.600e-08   1.127 0.260720
## vaccination.rate  4.735e-01  3.993e-01   1.186 0.236574
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 0.004849117)
##
##     Null deviance: 5.4174  on 353  degrees of freedom
## Residual deviance: 1.6039  on 337  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 6
```