Regresi Linear

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

#read library  
library(lmtest)

## Warning: package 'lmtest' was built under R version 4.0.5

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.0.3

##   
## Attaching package: 'zoo'

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

library(regclass)

## Warning: package 'regclass' was built under R version 4.0.5

## Loading required package: bestglm

## Warning: package 'bestglm' was built under R version 4.0.5

## Loading required package: leaps

## Warning: package 'leaps' was built under R version 4.0.3

## Loading required package: VGAM

## Warning: package 'VGAM' was built under R version 4.0.5

## Loading required package: stats4

## Loading required package: splines

##   
## Attaching package: 'VGAM'

## The following object is masked from 'package:lmtest':  
##   
## lrtest

## Loading required package: rpart

## Warning: package 'rpart' was built under R version 4.0.3

## Loading required package: randomForest

## Warning: package 'randomForest' was built under R version 4.0.3

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

## Important regclass change from 1.3:  
## All functions that had a . in the name now have an \_  
## all.correlations -> all\_correlations, cor.demo -> cor\_demo, etc.

library(olsrr)

## Warning: package 'olsrr' was built under R version 4.0.5

##   
## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':  
##   
## rivers

library(skedastic)

## Warning: package 'skedastic' was built under R version 4.0.5

#loading data  
strfile="E:\\xampp\\htdocs\\ANN COVID-19\\ALL\_FINAL.csv"  
data\_covid\_IHSG=read.csv(file = strfile)

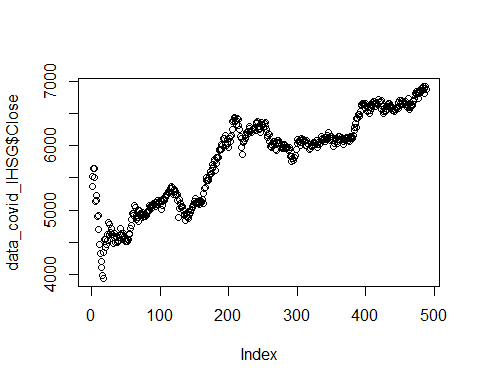
#cleaning data  
head(data\_covid\_IHSG)

## Date Close new\_cases new\_deaths pneumonia\_mean reproduction\_rate  
## 1 2020-03-02 5361.246 2 0 0.9586 0  
## 2 2020-03-03 5518.628 0 0 0.9597 0  
## 3 2020-03-04 5650.136 0 0 0.9602 0  
## 4 2020-03-05 5638.130 0 0 0.9609 0  
## 5 2020-03-06 5498.540 2 0 0.9622 0  
## 6 2020-03-09 5136.809 13 0 0.9662 0  
## new\_tests positive\_rate tests\_per\_case stringency\_index mask\_use\_mean  
## 1 0 0 0 28.70 0  
## 2 0 0 0 37.04 0  
## 3 0 0 0 37.04 0  
## 4 0 0 0 37.04 0  
## 5 0 0 0 37.04 0  
## 6 0 0 0 37.04 0  
## mobility\_mean total\_boosters total\_vaccinations new\_vaccinations  
## 1 2.4519 0 0 0  
## 2 2.7448 0 0 0  
## 3 2.9966 0 0 0  
## 4 3.1960 0 0 0  
## 5 3.3219 0 0 0  
## 6 2.9423 0 0 0  
## cumulative\_all\_vaccinated cumulative\_all\_fully\_vaccinated  
## 1 0 0  
## 2 0 0  
## 3 0 0  
## 4 0 0  
## 5 0 0  
## 6 0 0  
## cumulative\_all\_effectively\_vaccinated infection\_fatality infection\_detection  
## 1 0 0 0.0000  
## 2 0 0 0.0000  
## 3 0 0 0.0000  
## 4 0 0 0.0000  
## 5 0 0 0.0006  
## 6 0 0 0.0012  
## infection\_hospitalization  
## 1 0.0224  
## 2 0.0243  
## 3 0.0232  
## 4 0.0201  
## 5 0.0144  
## 6 0.0116

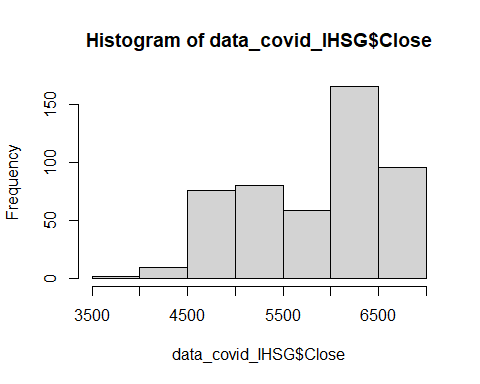
shapiro.test(data\_covid\_IHSG$Close)

##   
## Shapiro-Wilk normality test  
##   
## data: data\_covid\_IHSG$Close  
## W = 0.93134, p-value = 3.536e-14

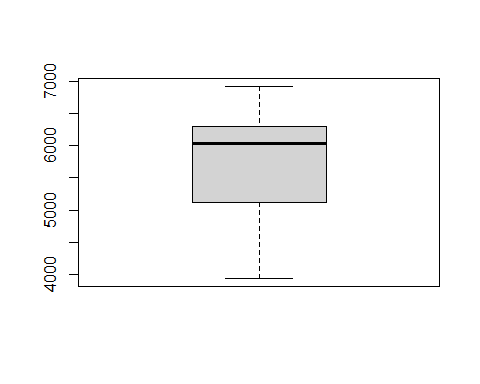
plot(data\_covid\_IHSG$Close)



hist(data\_covid\_IHSG$Close)



boxplot(data\_covid\_IHSG$Close)



LMfom=Close~.-Date-total\_boosters-positive\_rate-tests\_per\_case-stringency\_index-new\_vaccinations-cumulative\_all\_fully\_vaccinated-cumulative\_all\_effectively\_vaccinated-infection\_fatality-new\_tests

#formula and regression  
linearmodel<-lm(LMfom,data = data\_covid\_IHSG)

#result  
summary(linearmodel)

##   
## Call:  
## lm(formula = LMfom, data = data\_covid\_IHSG)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -455.2 -100.8 8.3 113.0 471.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.289e+03 1.759e+02 41.430 < 2e-16 \*\*\*  
## new\_cases 3.249e-03 1.127e-03 2.883 0.00412 \*\*   
## new\_deaths 2.695e-01 3.932e-02 6.856 2.21e-11 \*\*\*  
## pneumonia\_mean -2.959e+03 2.131e+02 -13.887 < 2e-16 \*\*\*  
## reproduction\_rate -1.528e+02 3.107e+01 -4.918 1.20e-06 \*\*\*  
## mask\_use\_mean 7.247e+02 1.245e+02 5.822 1.07e-08 \*\*\*  
## mobility\_mean 1.592e+01 1.433e+00 11.113 < 2e-16 \*\*\*  
## total\_vaccinations -1.478e-06 2.378e-07 -6.214 1.13e-09 \*\*\*  
## cumulative\_all\_vaccinated 7.458e-06 3.928e-07 18.985 < 2e-16 \*\*\*  
## infection\_detection 3.100e+04 2.356e+03 13.158 < 2e-16 \*\*\*  
## infection\_hospitalization 4.300e+04 5.691e+03 7.556 2.14e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 176.2 on 477 degrees of freedom  
## Multiple R-squared: 0.9379, Adjusted R-squared: 0.9366   
## F-statistic: 720.7 on 10 and 477 DF, p-value: < 2.2e-16

sqrt(mean(residuals.lm(linearmodel)^2))

## [1] 174.1552

#assumption check  
ols\_test\_normality(linearmodel)

## -----------------------------------------------  
## Test Statistic pvalue   
## -----------------------------------------------  
## Shapiro-Wilk 0.9889 9e-04   
## Kolmogorov-Smirnov 0.0509 0.1600   
## Cramer-von Mises 40.7111 0.0000   
## Anderson-Darling 1.6243 4e-04   
## -----------------------------------------------

lmtest::bptest(linearmodel)

##   
## studentized Breusch-Pagan test  
##   
## data: linearmodel  
## BP = 71.165, df = 10, p-value = 2.64e-11

lmtest::gqtest(linearmodel)

##   
## Goldfeld-Quandt test  
##   
## data: linearmodel  
## GQ = 0.24255, df1 = 233, df2 = 233, p-value = 1  
## alternative hypothesis: variance increases from segment 1 to 2

lmtest::hmctest(linearmodel)

##   
## Harrison-McCabe test  
##   
## data: linearmodel  
## HMC = 0.73395, p-value = 1

dwtest(linearmodel)

##   
## Durbin-Watson test  
##   
## data: linearmodel  
## DW = 0.21158, p-value < 2.2e-16  
## alternative hypothesis: true autocorrelation is greater than 0

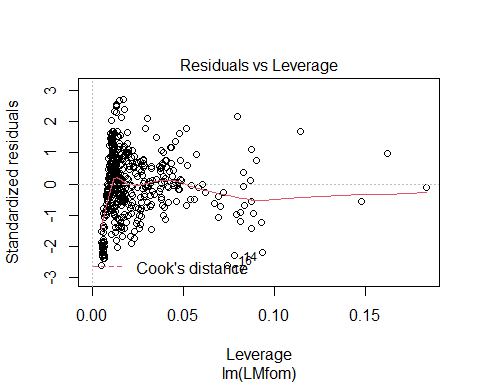
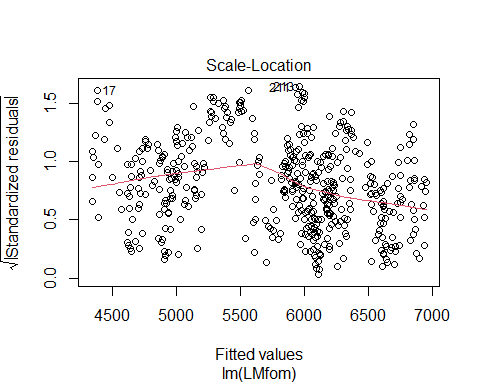
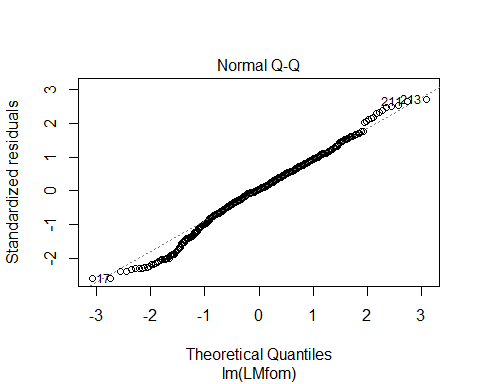
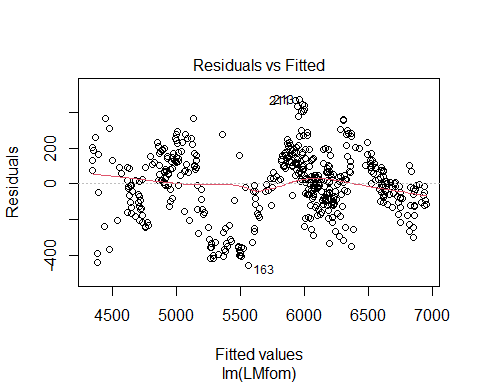
lmtest::bgtest(linearmodel)

##   
## Breusch-Godfrey test for serial correlation of order up to 1  
##   
## data: linearmodel  
## LM test = 391.02, df = 1, p-value < 2.2e-16

#correlated variables  
VIF(linearmodel)

## new\_cases new\_deaths pneumonia\_mean   
## 2.994805 3.012552 3.754501   
## reproduction\_rate mask\_use\_mean mobility\_mean   
## 1.907025 5.782850 2.997191   
## total\_vaccinations cumulative\_all\_vaccinated infection\_detection   
## 4.900090 9.140721 7.090305   
## infection\_hospitalization   
## 5.970331

plot(linearmodel)



lmauto=dwtest(linearmodel)  
p=(1-(lmauto$statistic/2))

Y\_varlm<-data\_covid\_IHSG$Close  
X\_varlm<-subset(data\_covid\_IHSG,select=-c(Date,total\_boosters,Close))

LMfomIt=Close~.-positive\_rate-tests\_per\_case-stringency\_index-new\_vaccinations-cumulative\_all\_fully\_vaccinated-cumulative\_all\_effectively\_vaccinated-infection\_fatality-new\_tests

#remove auto  
#iter=1  
Y\_varlm<-Y\_varlm[2:length(Y\_varlm)]-(p\*Y\_varlm[1:length(Y\_varlm)-1])  
X\_varlm<-X\_varlm[2:nrow(X\_varlm),]-(p\*X\_varlm[1:nrow(X\_varlm)-1,])  
data\_transformlm=data.frame("Close"=Y\_varlm,X\_varlm)  
  
lmFitIt<-lm(LMfomIt,data\_transformlm)  
lmautoIt=dwtest(lmFitIt)  
p=(1-(lmautoIt$statistic/2))  
lmautoIt

##   
## Durbin-Watson test  
##   
## data: lmFitIt  
## DW = 1.7881, p-value = 0.005169  
## alternative hypothesis: true autocorrelation is greater than 0

#iter=2  
Y\_varlm<-Y\_varlm[2:length(Y\_varlm)]-(p\*Y\_varlm[1:length(Y\_varlm)-1])  
X\_varlm<-X\_varlm[2:nrow(X\_varlm),]-(p\*X\_varlm[1:nrow(X\_varlm)-1,])  
data\_transformlm=data.frame("Close"=Y\_varlm,X\_varlm)  
  
lmFitIt<-lm(LMfomIt,data\_transformlm)  
lmautoIt=dwtest(lmFitIt)  
p=(1-(lmautoIt$statistic/2))  
lmautoIt

##   
## Durbin-Watson test  
##   
## data: lmFitIt  
## DW = 2.0027, p-value = 0.4379  
## alternative hypothesis: true autocorrelation is greater than 0

summary(lmFitIt)

##   
## Call:  
## lm(formula = LMfomIt, data = data\_transformlm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -266.508 -34.728 2.703 42.189 315.132   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7.011e+02 5.183e+01 13.527 < 2e-16 \*\*\*  
## new\_cases 3.693e-04 9.089e-04 0.406 0.684719   
## new\_deaths 4.640e-02 3.323e-02 1.396 0.163221   
## pneumonia\_mean -3.263e+03 5.330e+02 -6.123 1.93e-09 \*\*\*  
## reproduction\_rate -1.066e+02 3.234e+01 -3.297 0.001052 \*\*   
## mask\_use\_mean 1.514e+03 2.727e+02 5.551 4.73e-08 \*\*\*  
## mobility\_mean 9.423e+00 3.265e+00 2.886 0.004082 \*\*   
## total\_vaccinations -2.327e-07 2.518e-07 -0.924 0.355935   
## cumulative\_all\_vaccinated 7.168e-06 7.187e-07 9.973 < 2e-16 \*\*\*  
## infection\_detection 1.355e+04 3.788e+03 3.577 0.000384 \*\*\*  
## infection\_hospitalization 3.528e+04 9.231e+03 3.822 0.000150 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 69.05 on 475 degrees of freedom  
## Multiple R-squared: 0.4702, Adjusted R-squared: 0.4591   
## F-statistic: 42.16 on 10 and 475 DF, p-value: < 2.2e-16

sqrt(mean(residuals.lm(lmFitIt)^2))

## [1] 68.26369

LMfomIt=Close~.-positive\_rate-tests\_per\_case-stringency\_index-new\_vaccinations-cumulative\_all\_fully\_vaccinated-cumulative\_all\_effectively\_vaccinated-infection\_fatality-new\_tests-new\_cases-new\_deaths-total\_vaccinations

lmFitIt<-lm(LMfomIt,data\_transformlm)  
summary(lmFitIt)

##   
## Call:  
## lm(formula = LMfomIt, data = data\_transformlm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -266.045 -35.852 4.046 41.381 314.740   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.927e+02 5.114e+01 13.546 < 2e-16 \*\*\*  
## pneumonia\_mean -3.202e+03 5.271e+02 -6.076 2.52e-09 \*\*\*  
## reproduction\_rate -9.908e+01 3.153e+01 -3.143 0.001778 \*\*   
## mask\_use\_mean 1.544e+03 2.718e+02 5.682 2.32e-08 \*\*\*  
## mobility\_mean 8.555e+00 3.238e+00 2.642 0.008507 \*\*   
## cumulative\_all\_vaccinated 7.108e-06 6.770e-07 10.499 < 2e-16 \*\*\*  
## infection\_detection 1.384e+04 3.780e+03 3.661 0.000279 \*\*\*  
## infection\_hospitalization 3.314e+04 9.027e+03 3.671 0.000269 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 69.14 on 478 degrees of freedom  
## Multiple R-squared: 0.4655, Adjusted R-squared: 0.4577   
## F-statistic: 59.48 on 7 and 478 DF, p-value: < 2.2e-16

sqrt(mean(residuals.lm(lmFitIt)^2))

## [1] 68.56661

#assumption check  
ols\_test\_normality(lmFitIt)

## -----------------------------------------------  
## Test Statistic pvalue   
## -----------------------------------------------  
## Shapiro-Wilk 0.9823 0.0000   
## Kolmogorov-Smirnov 0.0536 0.1227   
## Cramer-von Mises 39.9864 0.0000   
## Anderson-Darling 1.7775 1e-04   
## -----------------------------------------------

lmtest::bptest(lmFitIt)

##   
## studentized Breusch-Pagan test  
##   
## data: lmFitIt  
## BP = 96.463, df = 7, p-value < 2.2e-16

lmtest::gqtest(lmFitIt)

##   
## Goldfeld-Quandt test  
##   
## data: lmFitIt  
## GQ = 0.35822, df1 = 235, df2 = 235, p-value = 1  
## alternative hypothesis: variance increases from segment 1 to 2

lmtest::hmctest(lmFitIt)

##   
## Harrison-McCabe test  
##   
## data: lmFitIt  
## HMC = 0.70035, p-value = 1

glejser(lmFitIt)

## # A tibble: 1 x 4  
## statistic p.value parameter alternative  
## <dbl> <dbl> <dbl> <chr>   
## 1 91.8 5.37e-17 7 greater

dwtest(lmFitIt)

##   
## Durbin-Watson test  
##   
## data: lmFitIt  
## DW = 1.9751, p-value = 0.3132  
## alternative hypothesis: true autocorrelation is greater than 0

VIF(lmFitIt)

## pneumonia\_mean reproduction\_rate mask\_use\_mean   
## 1.474212 1.072300 1.294254   
## mobility\_mean cumulative\_all\_vaccinated infection\_detection   
## 1.496115 1.736567 1.659099   
## infection\_hospitalization   
## 1.919576

plot(lmFitIt)

