Cau\_lenh\_tren\_phan\_mem.R

Administrator

2021-11-17

setwd("C:/Users/Administrator/Desktop/du\_an\_KTL/du\_lieu\_chinh")  
getwd()

## [1] "C:/Users/Administrator/Desktop/du\_an\_KTL/du\_lieu\_chinh"

#Doc file du lieu:  
library(readxl)   
test = read\_excel("du\_lieu\_chinh.xlsx", sheet = 1)  
  
#Thong ke mo ta:  
#Thong ke cac bien dinh tinh:  
library(table1)

##   
## Attaching package: 'table1'

## The following objects are masked from 'package:base':  
##   
## units, units<-

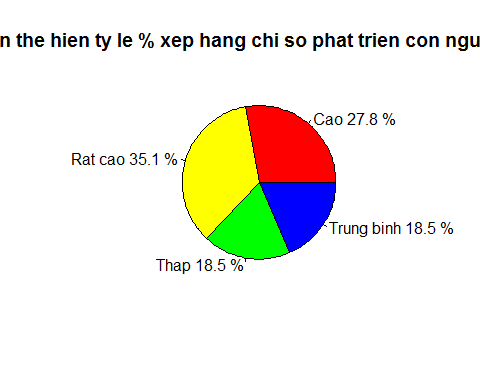
table1 (~total\_deaths + total\_vaccinations + population\_2020 +aged\_70\_older+   
 gdp\_per\_capita + diabetes\_prevalence + low\_HDI + medium\_HDI + high\_HDI, data = test)

|  | Overall (N=151) |
| --- | --- |
| **total\_deaths** |  |
| Mean (SD) | 31600 (92400) |
| Median [Min, Max] | 4080 [38.0, 724000] |
| **total\_vaccinations** |  |
| Mean (SD) | 28400000 (95500000) |
| Median [Min, Max] | 4220000 [35100, 973000000] |
| **population\_2020** |  |
| Mean (SD) | 38600000 (123000000) |
| Median [Min, Max] | 10100000 [98700, 1390000000] |
| **aged\_70\_older** |  |
| Mean (SD) | 5.53 (4.22) |
| Median [Min, Max] | 3.87 [0.617, 16.2] |
| **gdp\_per\_capita** |  |
| Mean (SD) | 18400 (19700) |
| Median [Min, Max] | 11800 [661, 117000] |
| **diabetes\_prevalence** |  |
| Mean (SD) | 7.57 (3.80) |
| Median [Min, Max] | 7.11 [0.990, 17.7] |
| **low\_HDI** |  |
| Mean (SD) | 0.185 (0.390) |
| Median [Min, Max] | 0 [0, 1.00] |
| **medium\_HDI** |  |
| Mean (SD) | 0.185 (0.390) |
| Median [Min, Max] | 0 [0, 1.00] |
| **high\_HDI** |  |
| Mean (SD) | 0.278 (0.450) |
| Median [Min, Max] | 0 [0, 1.00] |

#Phu luc:  
#Thong ke bien gia:  
#Ma hoa cac quoc gia:  
test$phat\_trien[test$HDI == 1] = "Thap"

## Warning: Unknown or uninitialised column: `phat\_trien`.

test$phat\_trien[test$HDI == 2] = "Trung binh"  
test$phat\_trien[test$HDI == 3] = "Cao"  
test$phat\_trien[test$HDI == 4] = "Rat cao"  
#Ve bieu do:  
dfa = table (test$phat\_trien)  
options (digits = 3)  
dfb = prop.table(dfa)  
dfc = paste(names(dfa), round(dfb\*100,digits = 1), "%")  
pie(dfa, labels = dfc, col = c("red","yellow","green","blue"), main = "Bieu do tron the hien ty le % xep hang chi so phat trien con nguoi cac quoc gia")



#Thong ke mo ta bien dinh tinh:  
library(gmodels)  
CrossTable (test$low\_HDI)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 151   
##   
##   
## | 0 | 1 |   
## |-----------|-----------|  
## | 123 | 28 |   
## | 0.815 | 0.185 |   
## |-----------|-----------|  
##   
##   
##   
##

CrossTable(test$medium\_HDI)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 151   
##   
##   
## | 0 | 1 |   
## |-----------|-----------|  
## | 123 | 28 |   
## | 0.815 | 0.185 |   
## |-----------|-----------|  
##   
##   
##   
##

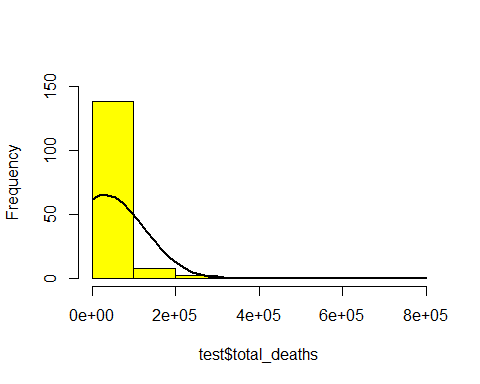
CrossTable (test$high\_HDI)

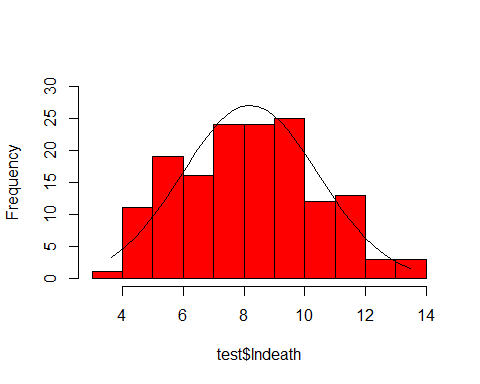
##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 151   
##   
##   
## | 0 | 1 |   
## |-----------|-----------|  
## | 109 | 42 |   
## | 0.722 | 0.278 |   
## |-----------|-----------|  
##   
##   
##   
##

CrossTable (test$very\_high\_HDI)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 151   
##   
##   
## | 0 | 1 |   
## |-----------|-----------|  
## | 98 | 53 |   
## | 0.649 | 0.351 |   
## |-----------|-----------|  
##   
##   
##   
##

#Ve do thi histogram cho bien total\_death:  
a = hist(test$total\_deaths, ylim = c(0,150), col = "yellow", main = "")  
b <- seq(min(test$total\_deaths), max(test$total\_deaths), length = 40)   
c <- dnorm(b, mean = mean(test$total\_deaths), sd = sd(test$total\_deaths))   
c <- c \* diff(a$mids[1:2]) \* length(test$total\_deaths)   
lines(b, c, col = "black", lwd = 2)



#Ve do thi histogram cho bien lndeath:  
h = hist (test$lndeath, breaks=10, ylim=c(0, 30), col = "red", main = "")  
xfit <- seq(min(test$lndeath), max(test$lndeath), length = 40)   
yfit <- dnorm(xfit, mean = mean(test$lndeath), sd = sd(test$lndeath))   
yfit <- yfit \* diff(h$mids[1:2]) \* length(test$lndeath)   
lines(xfit, yfit, col = "black", lwd = 1)

#He so tuong quan Pearson giua cac bien:  
cor (test$high\_HDI, test$lndeath)

## [1] 0.144

cor (test$high\_HDI, test$ln\_vaccine)

## [1] 0.0366

cor (test$high\_HDI, test$ln\_pop)

## [1] -0.0945

cor (test$high\_HDI, test$aged\_70\_older)

## [1] -0.101

cor (test$high\_HDI, test$gdp\_per\_capita)

## [1] -0.173

cor (test$high\_HDI, test$diabetes\_prevalence)

## [1] 0.267

cor (test$high\_HDI, test$low\_HDI)

## [1] -0.296

cor (test$high\_HDI, test$medium\_HDI)

## [1] -0.296

cor (test$medium\_HDI, test$lndeath)

## [1] -0.112

cor (test$medium\_HDI, test$ln\_vaccine)

## [1] -0.0586

cor (test$medium\_HDI, test$ln\_pop)

## [1] 0.0447

cor (test$medium\_HDI, test$aged\_70\_older)

## [1] -0.335

cor (test$medium\_HDI, test$gdp\_per\_capita)

## [1] -0.31

cor (test$medium\_HDI, test$diabetes\_prevalence)

## [1] -0.0746

cor (test$medium\_HDI, test$low\_HDI)

## [1] -0.228

cor (test$low\_HDI, test$lndeath)

## [1] -0.384

cor (test$low\_HDI, test$ln\_vaccine)

## [1] -0.387

cor (test$low\_HDI, test$ln\_pop)

## [1] 0.129

cor (test$low\_HDI, test$aged\_70\_older)

## [1] -0.426

cor (test$low\_HDI, test$gdp\_per\_capita)

## [1] -0.4

cor (test$low\_HDI, test$diabetes\_prevalence)

## [1] -0.372

cor (test$diabetes\_prevalence, test$lndeath)

## [1] 0.00702

cor (test$diabetes\_prevalence, test$ln\_vaccine)

## [1] 0.0225

cor (test$diabetes\_prevalence, test$ln\_pop)

## [1] -0.181

cor (test$diabetes\_prevalence, test$aged\_70\_older)

## [1] -0.0692

cor (test$diabetes\_prevalence, test$gdp\_per\_capita)

## [1] 0.245

cor (test$gdp\_per\_capita, test$lndeath)

## [1] 0.134

cor (test$gdp\_per\_capita, test$ln\_vaccine)

## [1] 0.27

cor (test$gdp\_per\_capita, test$ln\_pop)

## [1] -0.125

cor (test$gdp\_per\_capita, test$aged\_70\_older)

## [1] 0.483

cor (test$aged\_70\_older, test$lndeath)

## [1] 0.399

cor (test$aged\_70\_older, test$ln\_vaccine)

## [1] 0.346

cor (test$aged\_70\_older, test$ln\_pop)

## [1] -0.0371

cor (test$ln\_pop, test$lndeath)

## [1] 0.709

cor (test$ln\_pop, test$ln\_vaccine)

## [1] 0.747

cor (test$ln\_vaccine, test$lndeath)

## [1] 0.891

#Chay hoi quy:  
df1 = lm (data=test, lndeath ~ ln\_vaccine + ln\_pop + aged\_70\_older + gdp\_per\_capita  
 + diabetes\_prevalence + low\_HDI + medium\_HDI + high\_HDI)  
  
#Rut gon va xuat ket qua:  
library(jtools)

library(officer)

##   
## Attaching package: 'officer'

## The following object is masked from 'package:readxl':  
##   
## read\_xlsx

library(flextable)  
set\_summ\_defaults(digits = 3, model.info = TRUE, model.fit = TRUE, confint = TRUE, vifs = TRUE, pvals = TRUE, robust = TRUE)  
summ (df1)

Standard errors: Robust, type = HC3

export\_summs(df1, statistics = "all", robust = TRUE, scale = TRUE, to.file = "docx", file.name = "ket\_qua\_chay\_tren\_R.docx")

## Registered S3 methods overwritten by 'broom':  
## method from   
## tidy.glht jtools  
## tidy.summary.glht jtools

|  |  |
| --- | --- |
|  | Model 1 |
| (Intercept) | 8.704 \*\*\* |
|  | (0.222) |
| ln\_vaccine | 1.098 \*\*\* |
|  | (0.181) |
| ln\_pop | 0.815 \*\*\* |
|  | (0.164) |
| aged\_70\_older | 0.338 \*\*\* |
|  | (0.117) |
| gdp\_per\_capita | -0.448 \*\* |
|  | (0.109) |
| diabetes\_prevalence | -0.005 |
|  | (0.084) |
| low\_HDI | -1.698 \*\*\* |
|  | (0.472) |
| medium\_HDI | -1.022 \*\* |
|  | (0.368) |
| high\_HDI | 0.001 |
|  | (0.270) |
| nobs | 151 |
| r.squared | 0.868 |
| adj.r.squared | 0.861 |
| sigma | 0.830 |
| statistic | 116.915 |
| p.value | 0.000 |
| df | 8.000 |
| logLik | -181.465 |
| AIC | 382.931 |
| BIC | 413.104 |
| deviance | 97.799 |
| df.residual | 142.000 |
| nobs.1 | 151.000 |
| All continuous predictors are mean-centered and scaled by 1 standard deviation. Standard errors are heteroskedasticity robust. \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05. | |

#Kiem dinh RESET xem mo hinh df1 co phu hop khong:  
#Kiem dinh voi y2, y3:  
library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

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

resettest(df1, power = 2:3 , type = "fitted")

##   
## RESET test  
##   
## data: df1  
## RESET = 0.4, df1 = 2, df2 = 140, p-value = 0.7

#kiem dinh voi y2, y3, y4:  
resettest(df1, power = 2:4 , type = "fitted")

##   
## RESET test  
##   
## data: df1  
## RESET = 0.4, df1 = 3, df2 = 139, p-value = 0.7

#Kiem dinh Fisher:  
library(car)

## Loading required package: carData

linearHypothesis(df1, c("ln\_vaccine=0","ln\_pop=0","aged\_70\_older=0","gdp\_per\_capita=0","diabetes\_prevalence=0","low\_HDI=0","medium\_HDI=0","high\_HDI=0"))

## Linear hypothesis test  
##   
## Hypothesis:  
## ln\_vaccine = 0  
## ln\_pop = 0  
## aged\_70\_older = 0  
## gdp\_per\_capita = 0  
## diabetes\_prevalence = 0  
## low\_HDI = 0  
## medium\_HDI = 0  
## high\_HDI = 0  
##   
## Model 1: restricted model  
## Model 2: lndeath ~ ln\_vaccine + ln\_pop + aged\_70\_older + gdp\_per\_capita +   
## diabetes\_prevalence + low\_HDI + medium\_HDI + high\_HDI  
##   
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 150 742   
## 2 142 98 8 644 117 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Kiem dinh phuong sai thay doi:  
library(lmtest)  
bptest(df1)

##   
## studentized Breusch-Pagan test  
##   
## data: df1  
## BP = 14, df = 8, p-value = 0.07

#Kiem dinh phan phoi chuan:  
test$residual = residuals(df1)  
shapiro.test(test$residual)

##   
## Shapiro-Wilk normality test  
##   
## data: test$residual  
## W = 1, p-value = 0.3

#Kiem dinh da cong tuyen:  
vif (df1)

## ln\_vaccine ln\_pop aged\_70\_older gdp\_per\_capita   
## 7.16 5.87 3.00 2.60   
## diabetes\_prevalence low\_HDI medium\_HDI high\_HDI   
## 1.52 7.37 4.48 3.21