

Final exam task script

Ganesh vekhande

14/01/2022

```
Nations_df=read.csv(file.choose())
```

```
str(Nations_df)
```

```
## 'data.frame': 111 obs. of 6 variables:
## $ code : int 2001 2002 2003 2004 2005 2006 2007 2009 2010 2012 ...
## $ country: chr "Algeria" "Angola" "Benin" "Botswana" ...
## $ year : int 2004 2004 2004 2004 2004 2004 2004 2004 2004 2004 ...
## $ regime : int 1 0 1 2 0 0 0 0 0 0 ...
## $ rgdppc : num 4994 1897 867 3997 937 ...
## $ plit : int 0 0 0 0 0 0 0 0 0 0 ...
```

```
head(Nations_df)
```

```
## code country year regime rgdppc plit
## 1 2001 Algeria 2004 1 4993.58 0
## 2 2002 Angola 2004 0 1897.00 0
## 3 2003 Benin 2004 1 867.00 0
## 4 2004 Botswana 2004 2 3997.00 0
## 5 2005 BurkinaF 2004 0 937.00 0
## 6 2006 Burundi 2004 0 631.00 0
```

Data has 111 observations and 6 variables

Q.1

```
mean_rgdppc=mean(Nations_df$rgdppc)
median_rgdppc=median(Nations_df$rgdppc)
print(paste( " mean of real gross domestic product per capita is" , mean_rgdppc))
```

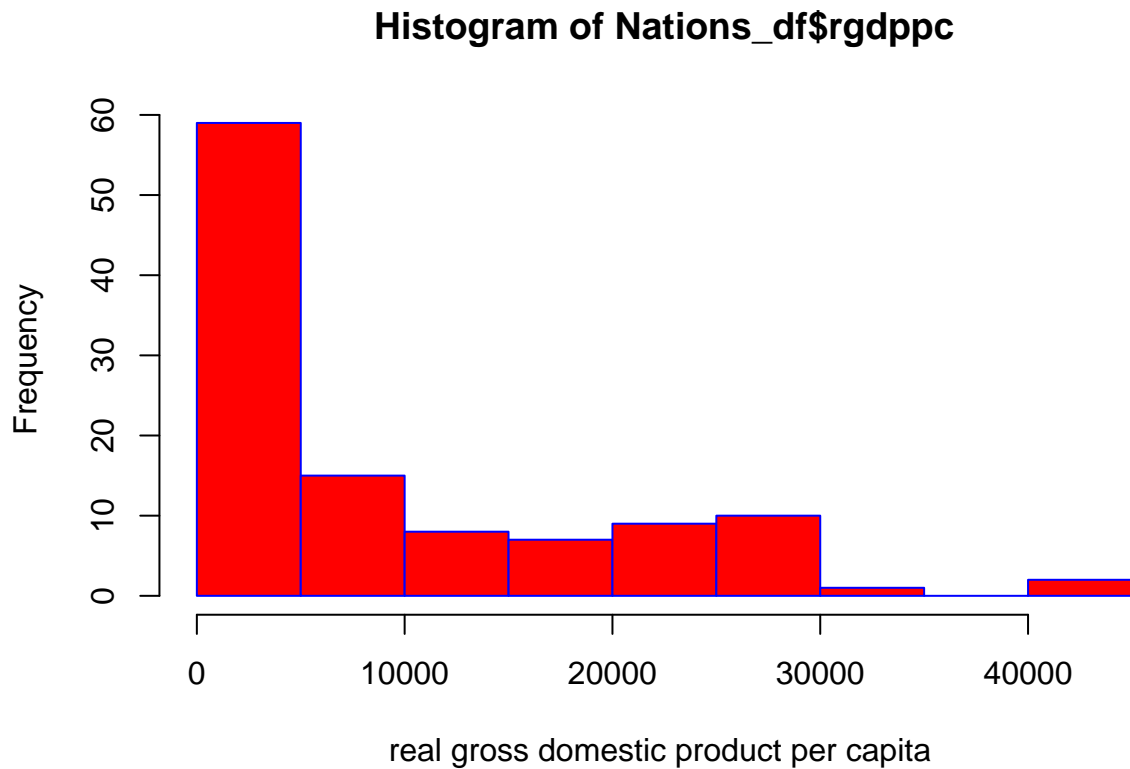
```
## [1] " mean of real gross domestic product per capita is 9314.95037585586"
```

```
print(paste( " median of real gross domestic product per capita is " , median_rgdppc))
```

```
## [1] " median of real gross domestic product per capita is 4613"
```

Q.2

```
h=hist(Nations_df$rgdppc,xlab = "real gross domestic product per capita ",col = "red",border = "blue")
```



```
h
```

```
## $breaks
## [1] 0 5000 10000 15000 20000 25000 30000 35000 40000 45000
##
## $counts
## [1] 59 15 8 7 9 10 1 0 2
##
## $density
## [1] 1.063063e-04 2.702703e-05 1.441441e-05 1.261261e-05 1.621622e-05
## [6] 1.801802e-05 1.801802e-06 0.000000e+00 3.603604e-06
##
## $mids
## [1] 2500 7500 12500 17500 22500 27500 32500 37500 42500
##
## $xname
## [1] "Nations_df$rgdppc"
##
## $equidist
## [1] TRUE
```

```
##
## attr(,"class")
## [1] "histogram"
```

```
mode <- h$mids[h$counts == max(h$counts)]
mode
```

```
## [1] 2500
```

from the graph we can clearly observe mean of real gross domestic product per capita is *9314.95037585586* median of real gross domestic product per capita is *4613* from hiatogrm mode of real gross domestic product per capita is *2500*

From here we can clearly see that maximum of countries has real gross doemstic product per capita as 2500 but since median Real gross domestic product per capita is 4613 and some of the countries have exceptionally high rgdppc which eventually affect the average rgdppc which is found higher as 9314.95

```
New_rgdppc=Nations_df$rgdppc[Nations_df$rgdppc <= 25000]
length(New_rgdppc)
```

so it will be good to remove countries which have rgdppc higher than 25000

```
## [1] 98
```

```
mean_new_rgdppc=mean(New_rgdppc)
median_new_rgdppc=median(New_rgdppc)
print(paste( " mean of real gross domestic product per capita is" ,mean_new_rgdppc ))
```

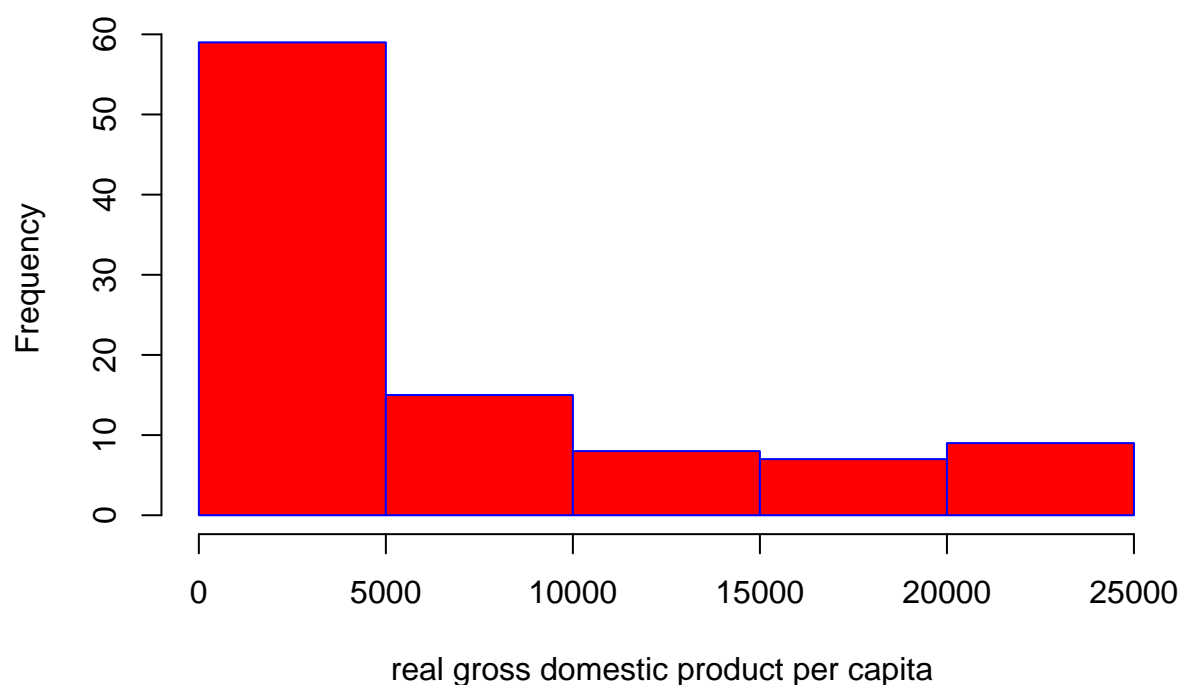
```
## [1] " mean of real gross domestic product per capita is 6597.92338489796"
```

```
print(paste( " median of real gross domestic product per capita is " , median_new_rgdppc))
```

```
## [1] " median of real gross domestic product per capita is 3877"
```

```
h1=hist(New_rgdppc,xlab = "real gross domestic product per capita ",col = "red",border = "blue")
```

Histogram of New_rgdpcc



```
h1
```

```
## $breaks
## [1] 0 5000 10000 15000 20000 25000
##
## $counts
## [1] 59 15 8 7 9
##
## $density
## [1] 1.204082e-04 3.061224e-05 1.632653e-05 1.428571e-05 1.836735e-05
##
## $mids
## [1] 2500 7500 12500 17500 22500
##
## $xname
## [1] "New_rgdpcc"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

```
model <- h1$mids[h1$counts == max(h1$counts)]
model
```

```
## [1] 2500
```

```
print(paste( " mean of real gross domestic product per capita is" ,mean_new_rgdppc ))
```

```
## [1] " mean of real gross domestic product per capita is 6597.92338489796"
```

```
print(paste( " median of real gross domestic product per capita is " , median_new_rgdppc))
```

```
## [1] " median of real gross domestic product per capita is 3877"
```

```
print(paste( " mode of real gross domestic product per capita is " , mode1 ))
```

```
## [1] " mode of real gross domestic product per capita is 2500"
```

Hence the problem is not completely but partly fixed that current data has no effect higher values of rgdppc

Q.3

```
cont_1=table(Nations_df$regime , Nations_df$plit)
cont_1
```

```
##
##      0  1
##  0 34  2
##  1 15  1
##  2 16 12
##  3  2 29
```

```
chisq.test(cont_1)
```

```
##
## Pearson's Chi-squared test
##
## data:  cont_1
## X-squared = 62.709, df = 3, p-value = 1.55e-13
```

as p value is $1.55e^{-13}$ which is less than 5 % level of significance hence we can reject null hypothesis that the regime is independent of plit

hence we conclude that regime type of nation depends on the literacy in the nation