MSE\_1740256\_Healthcare

Jeevan-1740256

31 January 2019

Health care or healthcare is the maintenance or improvement of health via the prevention, diagnosis, and treatment of disease, illness, injury, and other physical and mental impairments in human beings. Health indicators are quantifiable characteristics of a population which researchers use as supporting evidence for describing the health of a population. Typically, researchers will use a survey methodology to gather information about certain people, use statistics in an attempt to generalize the information collected to the entire population, then use the statistical analysis to make a statement about the health of the population. Health indicators are required in order to measure the health status of people and communities. They are often used by governments to guide health care policy. Analyse the given data and suggest a policy for government of India.

-To obtain and import a dataset in RStudio. -To find the descriptive statistics for various fields in the data obtained regarding the healthcare data to suggest a policy for the government of India. -To apply Linear Regression and fit the mean of the rural infant mortality rate to the graph of rural infant mortality rate and rural birth rate. -To find if the Rural birth rate and the infant mortality rate have any correlation. -To check whether there is any significant difference in the Total Birth Rate and Total Death Rate. -To visualize the data using various graphs such as line charts, kernel density plots.

setwd("C:/Users/Jeevan/Desktop/Christ University/R Studio")  
library(readr)

## Warning: package 'readr' was built under R version 3.5.2

BirthRate\_Data <- read\_csv("C:/Users/Jeevan/Desktop/Christ University/R Studio/BirthRate\_Data.csv")

## Warning: Missing column names filled in: 'X14' [14], 'X15' [15], 'X16' [16]

## Parsed with column specification:  
## cols(  
## Year = col\_double(),  
## RBR = col\_double(),  
## UBR = col\_double(),  
## TBR = col\_double(),  
## RDR = col\_double(),  
## UDR = col\_double(),  
## TDR = col\_double(),  
## RNGR = col\_double(),  
## UNGR = col\_double(),  
## TNGR = col\_double(),  
## RIMR = col\_double(),  
## UIMR = col\_double(),  
## TIMR = col\_double(),  
## X14 = col\_logical(),  
## X15 = col\_character(),  
## X16 = col\_character()  
## )

View(BirthRate\_Data)  
head(BirthRate\_Data, n = 5)

## # A tibble: 5 x 16  
## Year RBR UBR TBR RDR UDR TDR RNGR UNGR TNGR RIMR UIMR  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1994 30.5 23.1 28.7 10.1 6.7 9.3 20.4 16.4 19.4 80 52  
## 2 1995 30 22.7 28.3 9.8 6.6 9 20.2 16.1 19.3 80 48  
## 3 1996 29.3 21.6 27.5 9.7 6.5 9 19.6 15.1 18.5 77 46  
## 4 1997 28.9 21.5 27.2 9.6 6.5 8.9 19.3 15 18.3 77 45  
## 5 1998 28 21 26.5 9.7 6.6 9 18.3 14.5 17.5 77 45  
## # ... with 4 more variables: TIMR <dbl>, X14 <lgl>, X15 <chr>, X16 <chr>

summary(BirthRate\_Data)

## Year RBR UBR TBR   
## Min. :1994 Min. :22.10 Min. :17.00 Min. :20.40   
## 1st Qu.:2000 1st Qu.:23.50 1st Qu.:17.80 1st Qu.:21.95   
## Median :2005 Median :25.60 Median :19.00 Median :23.80   
## Mean :2005 Mean :25.74 Mean :19.38 Mean :24.10   
## 3rd Qu.:2010 3rd Qu.:27.60 3rd Qu.:20.75 3rd Qu.:25.95   
## Max. :2016 Max. :30.50 Max. :23.10 Max. :28.70   
## RDR UDR TDR RNGR   
## Min. : 6.900 Min. :5.400 Min. :6.400 Min. :15.20   
## 1st Qu.: 7.650 1st Qu.:5.750 1st Qu.:7.150 1st Qu.:15.80   
## Median : 8.100 Median :6.000 Median :7.500 Median :17.50   
## Mean : 8.435 Mean :6.017 Mean :7.804 Mean :17.33   
## 3rd Qu.: 9.350 3rd Qu.:6.300 3rd Qu.:8.600 3rd Qu.:18.25   
## Max. :10.100 Max. :6.700 Max. :9.300 Max. :20.40   
## UNGR TNGR RIMR UIMR   
## Min. :11.70 Min. :14.00 Min. :38.00 Min. :23.00   
## 1st Qu.:12.05 1st Qu.:14.80 1st Qu.:49.50 1st Qu.:30.00   
## Median :13.10 Median :16.30 Median :64.00 Median :39.00   
## Mean :13.38 Mean :16.29 Mean :61.83 Mean :37.35   
## 3rd Qu.:14.45 3rd Qu.:17.35 3rd Qu.:74.50 3rd Qu.:44.00   
## Max. :16.40 Max. :19.40 Max. :80.00 Max. :52.00   
## TIMR X14 X15 X16   
## Min. :34.0 Mode:logical Length:23 Length:23   
## 1st Qu.:45.5 NA's:23 Class :character Class :character   
## Median :58.0 Mode :character Mode :character   
## Mean :56.7   
## 3rd Qu.:69.0   
## Max. :74.0

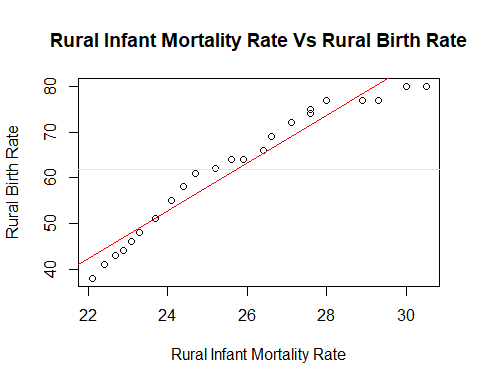
names(BirthRate\_Data)

## [1] "Year" "RBR" "UBR" "TBR" "RDR" "UDR" "TDR" "RNGR" "UNGR" "TNGR"  
## [11] "RIMR" "UIMR" "TIMR" "X14" "X15" "X16"

mean.rimr = mean(BirthRate\_Data$RIMR,na.rm = TRUE)  
mean.rimr

## [1] 61.82609

plot(BirthRate\_Data$RIMR~BirthRate\_Data$RBR,  
 data = BirthRate\_Data,  
 type = "p",  
 main = "Rural Infant Mortality Rate Vs Rural Birth Rate",  
 xlab = "Rural Infant Mortality Rate",  
 ylab = "Rural Birth Rate")  
abline(h=mean.rimr, col = "yellow")  
model1 = lm(BirthRate\_Data$RIMR~BirthRate\_Data$RBR, data = BirthRate\_Data)  
abline(model1, col = "red")



summary(model1)

##   
## Call:  
## lm(formula = BirthRate\_Data$RIMR ~ BirthRate\_Data$RBR, data = BirthRate\_Data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.7011 -2.9327 0.7405 2.9702 4.6310   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -72.804 7.190 -10.13 1.56e-09 \*\*\*  
## BirthRate\_Data$RBR 5.230 0.278 18.81 1.27e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.327 on 21 degrees of freedom  
## Multiple R-squared: 0.944, Adjusted R-squared: 0.9413   
## F-statistic: 353.9 on 1 and 21 DF, p-value: 1.271e-14

print("Correlation between Rural Birth Rate and Rural Infant Mortality Rate")

## [1] "Correlation between Rural Birth Rate and Rural Infant Mortality Rate"

cor.test(BirthRate\_Data$RBR,BirthRate\_Data$RIMR,  
 method = "pearson",  
 conf.level = 0.95)

##   
## Pearson's product-moment correlation  
##   
## data: BirthRate\_Data$RBR and BirthRate\_Data$RIMR  
## t = 18.812, df = 21, p-value = 1.271e-14  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.9330746 0.9880756  
## sample estimates:  
## cor   
## 0.9715888

var(BirthRate\_Data$TNGR, y = NULL)

## [1] 2.70083

var(BirthRate\_Data$TDR, y = NULL)

## [1] 0.7886166

print("T test")

## [1] "T test"

t.test(BirthRate\_Data$TNGR, BirthRate\_Data$TDR, var.equal = FALSE)

##   
## Welch Two Sample t-test  
##   
## data: BirthRate\_Data$TNGR and BirthRate\_Data$TDR  
## t = 21.789, df = 33.838, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 7.695245 9.278668  
## sample estimates:  
## mean of x mean of y   
## 16.291304 7.804348

chisq.test(BirthRate\_Data$RNGR,BirthRate\_Data$RDR)

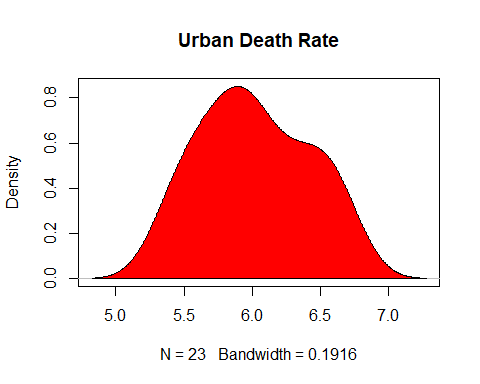
## Warning in chisq.test(BirthRate\_Data$RNGR, BirthRate\_Data$RDR): Chi-squared  
## approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: BirthRate\_Data$RNGR and BirthRate\_Data$RDR  
## X-squared = 310.5, df = 306, p-value = 0.4176

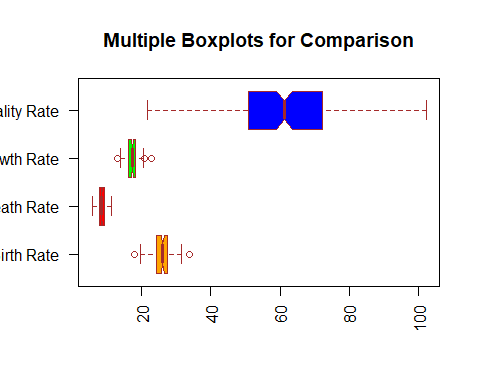
d = density(BirthRate\_Data$UDR)  
d

##   
## Call:  
## density.default(x = BirthRate\_Data$UDR)  
##   
## Data: BirthRate\_Data$UDR (23 obs.); Bandwidth 'bw' = 0.1916  
##   
## x y   
## Min. :4.825 Min. :0.001442   
## 1st Qu.:5.438 1st Qu.:0.081941   
## Median :6.050 Median :0.455118   
## Mean :6.050 Mean :0.407796   
## 3rd Qu.:6.662 3rd Qu.:0.662360   
## Max. :7.275 Max. :0.851094

plot(d, main = "Urban Death Rate")  
polygon(d,col="red",border="black")



rbr = BirthRate\_Data$RBR  
rdr = BirthRate\_Data$RDR  
rngr = BirthRate\_Data$RNGR  
rimr = BirthRate\_Data$RIMR  
rbr\_norm<-rnorm(200, mean = mean(rbr), sd = sd(rbr))  
rdr\_norm<-rnorm(200, mean = mean(rdr),sd = sd(rdr))  
rngr\_norm<-rnorm(200, mean = mean(rngr),sd = sd(rngr))  
rimr\_norm<-rnorm(200, mean = mean(rimr),sd = sd(rimr))  
boxplot(rbr\_norm, rdr\_norm, rngr\_norm, rimr\_norm, main="Multiple Boxplots for Comparison",  
 names = c("Rural Birth Rate","Rural Death Rate","Rural Natural Growth Rate","Rural Infant Mortality Rate"),  
 las=2,col = c("orange","red","green","blue"),border = "brown",horizontal = TRUE,notch = TRUE)



DESCRIPTIVE STATISTICS From the descriptive statistics thus obtained for the Rural Birth Rate, Urban Birth Rate, Total Birth Rate, Rural Death Rate, Urban Death Rate, Total Death Rate, Rural Natural Growth Rate, Urban Natural Growth Rate, Total Natural Growth Rate, Rural Infant Mortality Rate, Urban Infant Mortality Rate and the Total Infant Mortality Rate we see that the mean of Rural Infant Mortality Rate is the maximum.

CORRELATION Using the method Karl Pearson’s Product Moment Correlation, we see that the p-value 1.271e-14 which is less than 0.05 and therefore there is a strong correlation between the rural birth rate and the rural infant mortality rate.

T-TEST H0: There is no significant difference between the means of total natural growth rate and total death rate. H1: There is a significant difference between the means of total natural growth rate and total death rate..

From the results so obtained, we see that the p-value < 2.2e-16 which is less than the level of significance i.e. 0.05 due to which we reject the null hypothesis. Hence, we accept the alternative hypothesis that there is a significant difference between the means of the growth rates and the death rates.

CHI-SQUARE TESTS H0: The means of the total natural growth Rate and and total death rate are independent of each other. H1: The means of the total natural growth Rate and and total death rate are not independent of each other. From the result obtained, we see that the p-value = 0.4176 which is greater than the level of significance = 0.05 hence we accept the null hypothesis that they are independent.

VISUALISATION USING GRAPHS A linear regression model has been fitted to the graph between the rural infant mortality rate and rural birth rate. A kernel density plot of the urban death rate has been plotted which tells us that it is normally distributed. Multiple Boxplots have been used to compare all the rural factors.