Assignment 4

Reshmi

2022-09-02

#PROBLEM 1  
library(ggpubr)

## Loading required package: ggplot2

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

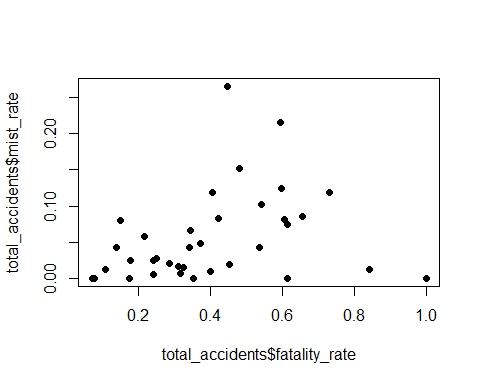
df <- read.csv('road\_accidents\_india\_2016.csv', row.names=1)  
acc\_cols <- grep("Total.Accidents$", colnames(df), ignore.case=T, value=TRUE)  
total\_accidents <- data.frame(state..ut=df$State..UT,total\_acc=rowSums(df[,c(acc\_cols)],na.rm=TRUE))  
print.data.frame(total\_accidents)

## state..ut total\_acc  
## 0 Andhra Pradesh 24888  
## 1 Arunachal Pradesh 249  
## 2 Assam 7435  
## 3 Bihar 8222  
## 4 Chhattisgarh 13580  
## 5 Goa 4304  
## 6 Gujarat 21859  
## 7 Haryana 11234  
## 8 Himachal Pradesh 3168  
## 9 Jammu & Kashmir 5501  
## 10 Jharkhand 4932  
## 11 Karnataka 44403  
## 12 Kerala 39420  
## 13 Madhya Pradesh 53972  
## 14 Maharashtra 39878  
## 15 Manipur 538  
## 16 Meghalaya 620  
## 17 Mizoram 83  
## 18 Nagaland 75  
## 19 Orissa 10532  
## 20 Punjab 6952  
## 21 Rajasthan 23066  
## 22 Sikkim 210  
## 23 Tamil Nadu 71431  
## 24 Telangana 22811  
## 25 Tripura 557  
## 26 Uttarakhand 1591  
## 27 Uttar Pradesh 35612  
## 28 West Bengal 13580  
## 29 A & N Islands 238  
## 30 Chandigarh 428  
## 31 D & N Haveli 70  
## 32 Daman & Diu 71  
## 33 Delhi 7375  
## 34 Lakshadweep 1  
## 35 Puducherry 1766

#PROBLEM 2  
death\_cols <- grep("Persons.Killed$", colnames(df), ignore.case=T, value=TRUE)  
  
total\_accidents$total\_deaths <- rowSums(df[ , c(death\_cols)])  
total\_accidents$fatality\_rate <- total\_accidents$total\_deaths / total\_accidents$total\_acc  
total\_accidents$mist\_rate <- df$Mist..Foggy...Total.Accidents / total\_accidents$total\_acc  
  
print.data.frame(total\_accidents)

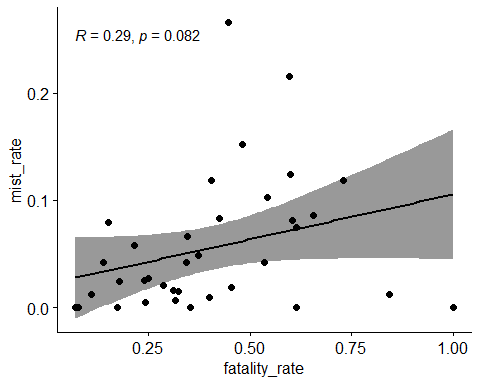
## state..ut total\_acc total\_deaths fatality\_rate mist\_rate  
## 0 Andhra Pradesh 24888 8541 0.34317743 0.042229187  
## 1 Arunachal Pradesh 249 149 0.59839357 0.124497992  
## 2 Assam 7435 2572 0.34593141 0.066039005  
## 3 Bihar 8222 4901 0.59608368 0.215154464  
## 4 Chhattisgarh 13580 3908 0.28777614 0.021207658  
## 5 Goa 4304 336 0.07806691 0.000000000  
## 6 Gujarat 21859 8136 0.37220367 0.048446864  
## 7 Haryana 11234 5024 0.44721382 0.265533203  
## 8 Himachal Pradesh 3168 1271 0.40119949 0.009785354  
## 9 Jammu & Kashmir 5501 958 0.17415015 0.000000000  
## 10 Jharkhand 4932 3027 0.61374696 0.074412003  
## 11 Karnataka 44403 11133 0.25072630 0.027520663  
## 12 Kerala 39420 4287 0.10875190 0.012683917  
## 13 Madhya Pradesh 53972 9646 0.17872230 0.024216260  
## 14 Maharashtra 39878 12935 0.32436431 0.014820202  
## 15 Manipur 538 81 0.15055762 0.079925651  
## 16 Meghalaya 620 150 0.24193548 0.004838710  
## 17 Mizoram 83 70 0.84337349 0.012048193  
## 18 Nagaland 75 46 0.61333333 0.000000000  
## 19 Orissa 10532 4463 0.42375617 0.083175085  
## 20 Punjab 6952 5077 0.73029344 0.118670886  
## 21 Rajasthan 23066 10465 0.45369808 0.018642157  
## 22 Sikkim 210 85 0.40476190 0.119047619  
## 23 Tamil Nadu 71431 17218 0.24104380 0.025353138  
## 24 Telangana 22811 7219 0.31647012 0.007233352  
## 25 Tripura 557 173 0.31059246 0.016157989  
## 26 Uttarakhand 1591 962 0.60465116 0.081081081  
## 27 Uttar Pradesh 35612 19320 0.54251376 0.102886667  
## 28 West Bengal 13580 6544 0.48188513 0.151767305  
## 29 A & N Islands 238 17 0.07142857 0.000000000  
## 30 Chandigarh 428 151 0.35280374 0.000000000  
## 31 D & N Haveli 70 46 0.65714286 0.085714286  
## 32 Daman & Diu 71 38 0.53521127 0.042253521  
## 33 Delhi 7375 1591 0.21572881 0.057627119  
## 34 Lakshadweep 1 1 1.00000000 0.000000000  
## 35 Puducherry 1766 244 0.13816535 0.042468856

plot(x=total\_accidents$fatality\_rate, y=total\_accidents$mist\_rate, pch=19)



rho <- cor(total\_accidents$fatality\_rate,total\_accidents$mist\_rate, method='pearson')  
  
corr\_test = cor.test(total\_accidents$fatality\_rate,total\_accidents$mist\_rate, method='pearson')  
  
ggscatter(total\_accidents, x='fatality\_rate',y='mist\_rate',add='reg.line',conf.int=TRUE, cor.coef=TRUE, cor.method = 'pearson')

## `geom\_smooth()` using formula 'y ~ x'

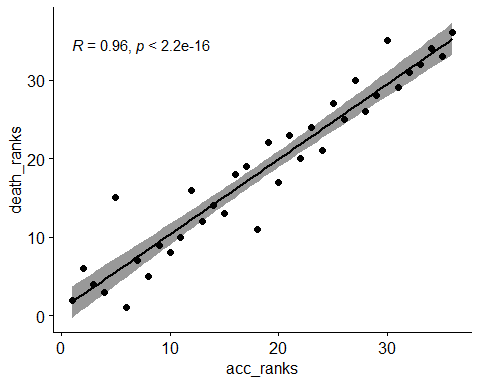


#PROBLEM 3  
total\_accidents$acc\_ranks <- rank(desc(total\_accidents$total\_acc),ties.method='random')  
total\_accidents$death\_ranks <- rank(desc(total\_accidents$total\_deaths),ties.method='random')  
  
rs <- cor(total\_accidents$acc\_ranks, total\_accidents$death\_ranks,method='spearman')  
  
print(cor.test(total\_accidents$acc\_ranks,total\_accidents$death\_ranks, method='spearman'))

##   
## Spearman's rank correlation rho  
##   
## data: total\_accidents$acc\_ranks and total\_accidents$death\_ranks  
## S = 326, p-value < 2.2e-16  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.9580438

ggscatter(total\_accidents, x='acc\_ranks', y='death\_ranks',add='reg.line', conf.int=TRUE,  
cor.coef=TRUE, cor.method = 'spearman')

## `geom\_smooth()` using formula 'y ~ x'



degrees <- nrow(total\_accidents) - 2  
t\_stat <- (rs - 0.2)/sqrt((1 - rs\*rs)/(nrow(total\_accidents) - 2))  
  
#Two-tailed test  
2 \* pt(q=t\_stat, df=degrees, lower.tail=FALSE)

## [1] 6.496376e-17

#PROBLEM 4  
total\_accidents$hail\_binary <- ifelse(df$Hail.Sleet...Total.Accidents > 0, 1, 0)  
total\_accidents$rain\_acc <- df$Rainy...Total.Accidents  
  
print.data.frame(total\_accidents[, c('state..ut', 'hail\_binary', 'rain\_acc')])

## state..ut hail\_binary rain\_acc  
## 0 Andhra Pradesh 1 1456  
## 1 Arunachal Pradesh 1 30  
## 2 Assam 1 528  
## 3 Bihar 0 939  
## 4 Chhattisgarh 0 1279  
## 5 Goa 0 529  
## 6 Gujarat 1 759  
## 7 Haryana 1 1656  
## 8 Himachal Pradesh 0 136  
## 9 Jammu & Kashmir 0 77  
## 10 Jharkhand 1 859  
## 11 Karnataka 1 3475  
## 12 Kerala 0 6902  
## 13 Madhya Pradesh 1 3931  
## 14 Maharashtra 0 1958  
## 15 Manipur 1 81  
## 16 Meghalaya 0 64  
## 17 Mizoram 0 0  
## 18 Nagaland 0 0  
## 19 Orissa 1 1637  
## 20 Punjab 0 402  
## 21 Rajasthan 1 475  
## 22 Sikkim 1 19  
## 23 Tamil Nadu 0 2893  
## 24 Telangana 0 237  
## 25 Tripura 0 30  
## 26 Uttarakhand 1 75  
## 27 Uttar Pradesh 1 3168  
## 28 West Bengal 1 2267  
## 29 A & N Islands 0 63  
## 30 Chandigarh 0 0  
## 31 D & N Haveli 0 15  
## 32 Daman & Diu 0 7  
## 33 Delhi 1 449  
## 34 Lakshadweep 0 0  
## 35 Puducherry 1 198

cor.test(total\_accidents$rain\_acc,  
total\_accidents$hail\_binary, method='pearson')

##   
## Pearson's product-moment correlation  
##   
## data: total\_accidents$rain\_acc and total\_accidents$hail\_binary  
## t = 0.84232, df = 34, p-value = 0.4055  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.1947090 0.4503544  
## sample estimates:  
## cor   
## 0.1429725

#PROBLEM 5  
total\_accidents$dust\_binary <- ifelse(df$Dust.Storm...Total.Accidents > 0, 1, 0)  
contingency\_table <- table(total\_accidents[, c('dust\_binary', 'hail\_binary')])  
  
library(psych)

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

phi(contingency\_table)

## [1] 0.62