Data Analytics

**IPL Data Analysis – Assignment - 2**

Name: - Ashish

Roll no.: - 17BCS006

|  |
| --- |
| > library(tidyr)  > library(dplyr)  > library(ggplot2)  > library(knitr)  >  > batsmen <- read.csv("C:/Users/Ashish/Desktop/R2/batsmen.csv")  > batsmen  >  > ##-------Descriptive statistics-------------##  > #---Minimum-------------#  > min(batsmen$Mat)  [1] 50  > min(batsmen$Inns)  [1] 50  > min(batsmen$NO)  [1] 2  > min(batsmen$Runs)  [1] 1441  >  > min(batsmen$HS)  [1] 48  > min(batsmen$Ave)  [1] 20.72  >  > #---Maximum-------------#  > max(batsmen$Mat)  [1] 193  > max(batsmen$Inns)  [1] 189  > max(batsmen$NO)  [1] 65  > max(batsmen$Runs)  [1] 5412  > max(batsmen$HS)  [1] 177  > max(batsmen$Ave)  [1] 43.17  > max(batsmen$BF)  [1] 4112  >  > #---Mean------------------#  > mean(batsmen$Mat)  [1] 113.96  > mean(batsmen$Inns)  [1] 105.98  > mean(batsmen$NO)  [1] 17.7  > mean(batsmen$Runs)  [1] 2722.6  > mean(batsmen$HS)  [1] 111.98  > mean(batsmen$Ave)  [1] 30.819  >  > #---Variance-------------#  > var(batsmen$Mat)  [1] 1687.713  > var(batsmen$Inns)  [1] 1487.163  > var(batsmen$NO)  [1] 169.0306  > var(batsmen$Runs)  [1] 1363186  > var(batsmen$HS)  [1] 697.7751  > var(batsmen$Ave)  [1] 33.42009  >  > #---Standard Deviation------------#  > sd(batsmen$Mat)  [1] 41.08178  > sd(batsmen$Inns)  [1] 38.56375  > sd(batsmen$NO)  [1] 13.00118  > sd(batsmen$Runs)  [1] 1167.556  > sd(batsmen$HS)  [1] 26.41543  >  > ##----------------Bowlers-------------------------##  >  > #---Taking bowlers data---------------#  > bowler<-read.csv("C:/Users/Ashish/Desktop/R2/bowler.csv")  >  > #---Find Minimum------------------------#  > min(bowler$Mat)  [1] 45  > min(bowler$Inns)  [1] 45  > min(bowler$Overs)  [1] 163.3  > min(bowler$Mdns)  [1] 0  > min(bowler$Runs)  [1] 1193  > min(bowler$Wkts)  [1] 47  >  > #---Find Maximum---------------#  > max(bowler$Mat)  [1] 170  > max(bowler$Inns)  [1] 157  > max(bowler$Overs)  [1] 562.2  > max(bowler$Mdns)  [1] 14  > max(bowler$Runs)  [1] 4072  > max(bowler$Wkts)  [1] 170  >  > #---Find Mean-----------------#  > mean(bowler$Mat)  [1] 91.59184  > mean(bowler$Inns)  [1] 88.02041  > mean(bowler$Overs)  [1] 305.8163  > mean(bowler$Mdns)  [1] 3.265306  > mean(bowler$Runs)  [1] 2363.02  > mean(bowler$Wkts)  [1] 88.16327  >  > #---Find Variance--------------#  > var(bowler$Mat)  [1] 1037.163  > var(bowler$Inns)  [1] 841.1037  > var(bowler$Overs)  [1] 10362.76  > var(bowler$Mdns)  [1] 8.365646  > var(bowler$Runs)  [1] 530165.7  > var(bowler$Wkts)  [1] 936.0978  >  > #---Find Standard Deviation--------------#  > sd(bowler$Mat)  [1] 32.20502  > sd(bowler$Inns)  [1] 29.00179  > sd(bowler$Overs)  [1] 101.7976  > sd(bowler$Mdns)  [1] 2.892343  > sd(bowler$Runs)  [1] 728.1248  > sd(bowler$Wkts)  [1] 30.59572  >  > #---Bar Plot of Batsmen vs Runs  > ggplot(batsmen) + geom\_bar(aes(Runs,Player, fill = Runs), stat = 'identity')  + coord\_flip()+ labs(title="Batsmen: Player vs Runs") + theme(axis.text.x =  element\_text(angle = 90))  >  >  >  > #---Bar Plot of Batsmen: Player vs Economy  > barplot(bowler$Econ,col = "red", pch = 19, main = "Batsmen: Player vs Economy",  xlab = "Player", ylab = "Economy")  >  > #---Histogram for Batsmen  > par(mfrow = c(5, 1))  > par(mar = rep(2, 4))  > hist(batsmen$Mat,main = "Histogram of Batsmen Matches")  > hist(batsmen$Inns,main = "Histogram of Batsmen Innings")  > hist(batsmen$Runs,main = "Histogram of Batsmen Runs")  > hist(batsmen$Ave,,main = "Histogram of Batsmen Average")  > hist(batsmen$SR,main = "Histogram of Batsmen Strike-Rate")  >  > #---Histogram for Bowlers Data  > par(mfrow = c(3, 3))  > par(mar = rep(2, 4))  > hist(bowler$Mat,main = "Histogram of Bowler Matches")  > hist(bowler$Inns,main = "Histogram of Bowler Innings")  > hist(bowler$Overs,main = "Histogram of Bowler overs")  > hist(bowler$Ave,,main = "Histogram of Bowler Average")  > hist(bowler$Econ,main = "Histogram of Bowler Economy")  > hist(bowler$Wkts,main = "Histogram of Bowlers Wickets")  >  > #---Detecting Outliers from Batsmen  > outlierKD(batsmen,Runs)  >  >  > #---Detecting Outliers from Batsman Matches  > outlierKD(batsman, mat)  >  >  > data <- read.csv("C:/Users/Ashish/Desktop/R2/batsmen.csv")  > data    > Bpoints<-((data$Runs\*10)+(data$Ave\*6)+(data$SR\*2)+(data$Fours\*9)+(data$Sixes\*11)  +(data$HF\*9))  > Bpoints  numeric(0)  >  > data1 <- read.csv("C:/Users/Ashish/Desktop/R2/bowler.csv")  > data1    > data1$bat = "Bpoints"  > data1[with(data1, order("bat")),]    >  > rum1 <- data1[order(data1$bat, decreasing = TRUE),]  >  > poin <- c((data$Runs\*10), (data$Ave\*6), (data$SR\*2), (data$Fours\*9),  (data$Sixes\*11), (data$HF\*9))  > barplot(poin)  >  > pairs(data[,2:7], pch=20, col="#FC4E07")    > d <- data[-1,-1]  > head(d)    > sapply(d,class)  > sapply(d, is.factor)  > cor(d[sapply(d, function(x) !is.factor(x))])  >  > d1 <- cor(d[sapply(d, function(x) !is.factor(x))])  > head(d1)  >  > d2 <- eigen(d1)$vectors  > head(d2)    >  > pc <- princomp(d1, cor = TRUE, scores = TRUE) #principal component  > pca <- prcomp(t(d1), scale = TRUE) #Principal Component Analysis  > pca$x    > plot(pca$x[,1], pca$x[,2])    > pca.var <- pca$sdev^2  > pca.var    > summary(pc)    > plot(pc)    > plot(pc,type="l")    > biplot(pc)    > dim(d)  [1] 49 15  > attributes(pc)  $names  [1] "sdev" "loadings" "center" "scale" "n.obs" "scores" "call"  $class  [1] "princomp"  > pc$loadings    > pc$scores    > pc$call  princomp(x = d1, cor = TRUE, scores = TRUE)  > pc$sdev    > pc$center    > pc$scale    > pc$n.obs  [1] 13  > str(pc)    > data2 <- read.csv("C:/Users/Ashish/Desktop/R2/bowler.csv")  > data2    > pairs(data2[,2:5], pch=20, col="#FC4E07")  > d3 <- data2[,-1]  > head(d3)  Player Span Mat Inns Overs Mdns Runs Wkts BBI Ave Econ SR X4 X5  1 SL Malinga 2009-2019 122 122 471.1 8 3366 170 5/13 19.80 7.14 16.6 6 1  2 A Mishra 2008-2019 147 147 516.5 6 3795 157 5/17 24.17 7.34 19.7 3 1  3 Harbhajan Singh 2008-2019 160 157 562.2 6 3967 150 5/18 26.44 7.05 22.4 1 1  4 PP Chawla 2008-2019 157 156 520.4 2 4072 150 4/17 27.14 7.82 20.8 2 0  5 DJ Bravo 2008-2019 134 131 430.5 2 3618 147 4/22 24.61 8.39 17.5 2 0  6 B Kumar 2011-2019 117 117 435.2 8 3154 133 5/19 23.71 7.24 19.6 2 1  >  > sapply(d3,class)  > sapply(d3, is.factor)  > cor(d3[sapply(d3, function(x) !is.factor(x))])  > #\*cov(d3)  > d4 <- cor(d3[sapply(d3, function(x) !is.factor(x))])  > head(d4)    > d5 <- eigen(d4)$vectors  > head(d5)    > pc1 <- princomp(d4, cor = TRUE, scores = TRUE)  > pca1 <- prcomp(t(d4), scale = TRUE)  > pca1$x    > plot(pca1$x[,1], pca1$x[,2])    > pca1.var <- pca1$sdev^2  > pca1.var    > summary(pc1)    > plot(pc1)    > plot(pc1,type="l")    > biplot(pc1)    > dim(d3)  [1] 49 14  > attributes(pc1)  $names  [1] "sdev" "loadings" "center" "scale" "n.obs" "scores" "call"  $class  [1] "princomp"  > pc1$loadings    > pc1$scores    > pc1$call  princomp(x = d4, cor = TRUE, scores = TRUE)  > pc1$sdev    > pc1$center    > pc1$scale    > pc1$n.obs  [1] 11  > str(pc1)  List of 7  $ sdev : Named num [1:11] 2.595 1.575 0.983 0.687 0.563 ...  ..- attr(\*, "names")= chr [1:11] "Comp.1" "Comp.2" "Comp.3" "Comp.4" ...  $ loadings: 'loadings' num [1:11, 1:11] 0.343 0.362 0.373 0.202 0.366 ...  ..- attr(\*, "dimnames")=List of 2  .. ..$ : chr [1:11] "Mat" "Inns" "Overs" "Mdns" ...  .. ..$ : chr [1:11] "Comp.1" "Comp.2" "Comp.3" "Comp.4" ...  $ center : Named num [1:11] 0.484 0.51 0.494 0.295 0.511 ...  ..- attr(\*, "names")= chr [1:11] "Mat" "Inns" "Overs" "Mdns" ...  $ scale : Named num [1:11] 0.386 0.432 0.461 0.301 0.424 ...  ..- attr(\*, "names")= chr [1:11] "Mat" "Inns" "Overs" "Mdns" ...  $ n.obs : int 11  $ scores : num [1:11, 1:11] 1.7813 2.3052 2.5496 0.0873 2.2948 ...  ..- attr(\*, "dimnames")=List of 2  .. ..$ : chr [1:11] "Mat" "Inns" "Overs" "Mdns" ...  .. ..$ : chr [1:11] "Comp.1" "Comp.2" "Comp.3" "Comp.4" ...  $ call : language princomp(x = d4, cor = TRUE, scores = TRUE)  - attr(\*, "class")= chr "princomp" |
|  |
| |  | | --- | | > | |