

NAME: 1) SRICHARAN ADAPA (21BCE6072)

2)AKASH.S (21BCE1705)

TOPIC: MOST RUNS IN INTERNATIONAL CRICKET

Table creation:

Rcode

Sl\_no=c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20)

name=c("Sachin Tendulkar","Kumar Sangakkara","Ricky Ponting","Mahela Jayawardene",

"Rahul Dravid","Virat Kohli","Brain Lara","AB de Villiers","Chris Gayle","Younis Khan",

"Joe Root", "MS Dhoni","Shahid Afridi","David Warner","Kane Williamson","Adam Gilchrist",

"Shoaib Malik","Babar Azam","Michael Hussey","Kevin Pietersen" )

nationality=c(0,1,2,1,0,0,3,4,3,5,6,0,5,2,7,2,5,5,2,6)

innings=c(782,666,668,725,605,527,521,484,551,491,405,526,508,407,381,429,429,252,324,342)

runs=c(34357,28016,27483,25957,24208,24130,22358,20014,19593,

17790,17604,17266,11196,16466,15889,15461,11867,11017,12398,13797)

average=c(48.52,46.77,45.95,39.15,45.41,53.62,46.28,48.11,37.97,39.88,49.03,44.96,23.92,43.10,46.45,38.94,33.90,50.53,49.00,44.07)

S\_rate=c(67.58,66.56,68.48,64.73,51.98,79.15,68.08,74.71,77.22,60.57,65.73,79.07,114.14,86.32,66.30,91.43,77.60,81.49,64.54,71.80)

Bat=data.frame(Sl\_no,name,nationality,innings,runs,average,S\_rate)

Bat

Bat$nationality=factor(Bat$nationality,labels=c("India","Sri Lanka","Australia","West Indies","South Africa","Pakistan","England","New Zealand"))

Output:



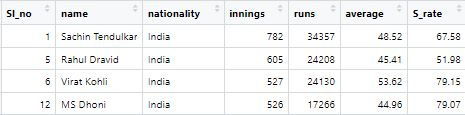
Tables based on nationality

1. India

Rcode:

Team\_Ind=subset(Bat,Bat$nationality=='India')

Output:

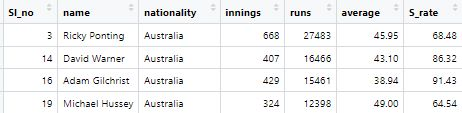


1. Australia

Rcode:

Team\_Aus=subset(Bat,Bat$nationality=='Australia')

Output:

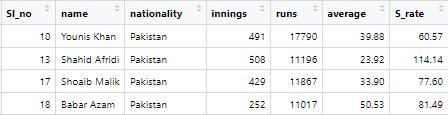


1. Pakistan

Rcode:

Team\_PAK=subset(Bat,Bat$nationality=='Pakistan')

Output:



Rcode for the rest of the teams:

Team\_SL=subset(Bat,Bat$nationality=='Sri Lanka')

Team\_NZ=subset(Bat,Bat$nationality=='New Zealand')

Team\_WI=subset(Bat,Bat$nationality=='West Indies')

Team\_RSA=subset(Bat,Bat$nationality=='South Africa')

Team\_Eng=subset(Bat,Bat$nationality=='England')

Measures of central tendency for runs of each team

1. India

Rcode:

summary(Team\_Ind$runs)

Output:



1. Australia

Rcode:

summary(Team\_Aus$runs)

Output:



1. Pakistan

Rcode:

summary(Team\_PAK$runs)

Output:



Rcode for the remaining teams:

summary(Team\_SL$runs)

summary(Team\_NZ$runs)

summary(Team\_WI$runs)

summary(Team\_RSA$runs)

summary(Team\_Eng$runs)

Correlation between the teams:

Between India and Australia , between Australia and Pakistan ,and between India and Pakistan

a1=Team\_Ind$runs

a2=Team\_Aus$runs

a3=Team\_PAK$runs

cor(a1,a2)

cor(a2,a3)

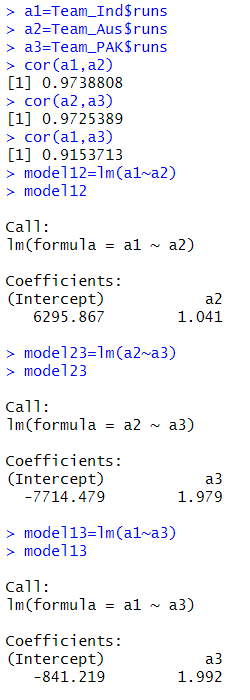
cor(a1,a3)

model12=lm(a1~a2)

model23=lm(a2~a3)

model13=lm(a1~a3)

Output:



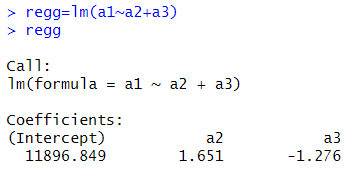
Multiple regression between the teams:

Rcode:

regg=lm(a1~a2+a3)

regg

Output:



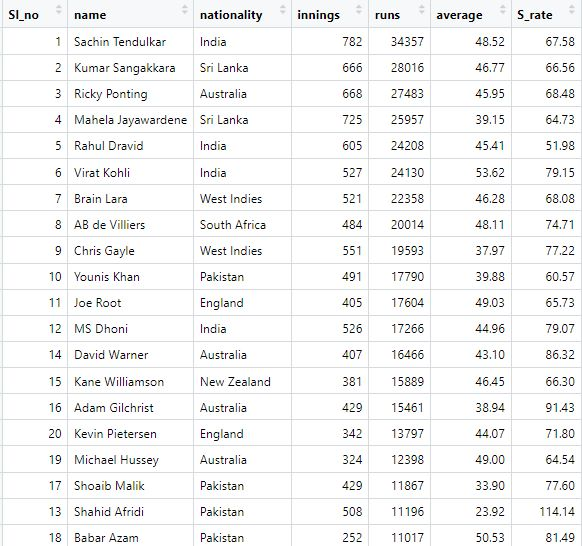
Sorting of players in decreasing order of the runs scored by them

Rcode:

df\_sorted =Bat[order(Bat$runs, decreasing = TRUE), ]

df\_sorted

Output:



Highest average by a batsman

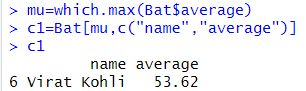
Rcode:

mu=which.max(Bat$average)

c1=Bat[mu,c("name","average")]

c1

Output:



Highest strike rate by a batsman

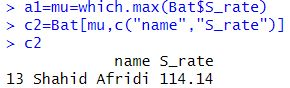
Rcode:

a1=mu=which.max(Bat$S\_rate)

c2=Bat[mu,c("name","S\_rate")]

c2

Output:

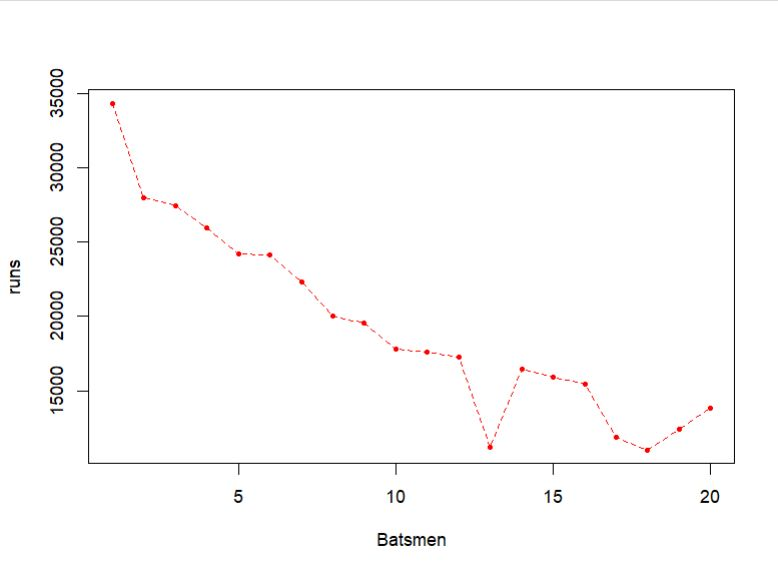


Plots

Rcode:

plot(runs,xlab = "Batsmen",type="o", pch=20,lty=2,col="red")

Output:



Rcode:

ind1=sum(Team\_Ind$runs)

aust=sum(Team\_Aus$runs)

paki=sum(Team\_PAK$runs)

sri=sum(Team\_SL$runs)

nz=sum(Team\_NZ$runs)

wi=sum(Team\_WI$runs)

sa=sum(Team\_RSA$runs)

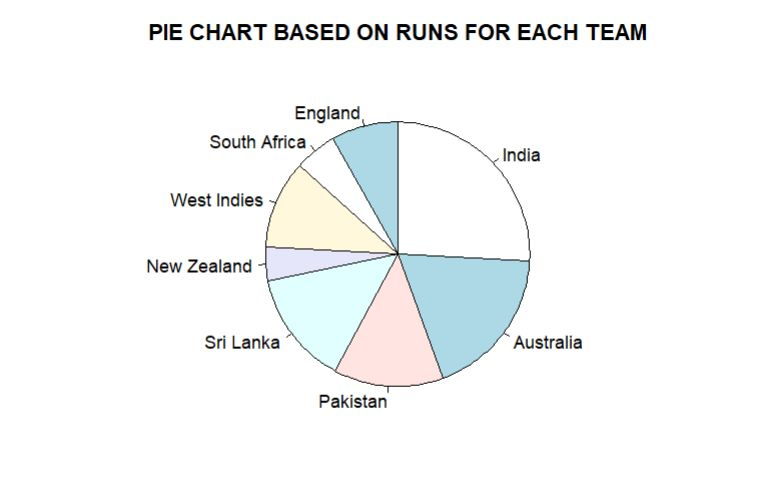
eng=sum(Team\_Eng$runs)

x2=c(ind1,aust,paki,sri,nz,wi,sa,eng)

x3=c("India","Australia","Pakistan","Sri Lanka","New Zealand","West Indies","South Africa","England")

pie(x2,labels=x3,clockwise = TRUE,main = "PIE CHART BASED ON RUNS FOR EACH TEAM")

Output:



T-test

Case 1:

H0: The mean runs scored is 19343.35

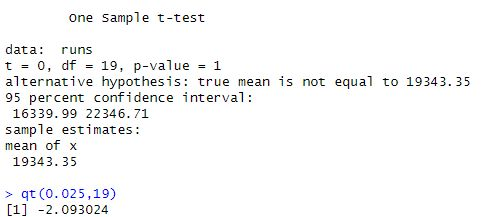
H1: The mean runs scored is not 19343.35

Rcode:

t.test(runs,alternative="two.sided",mu=19343.35)

qt(0.025,19)

Output:



Conclusion: Since t(cal)<t(critical), reject null hypothesis at 5% LOS. Thus, the mean runs scored is 19343.35.

Case 2:

H0: The mean runs scored is 9000

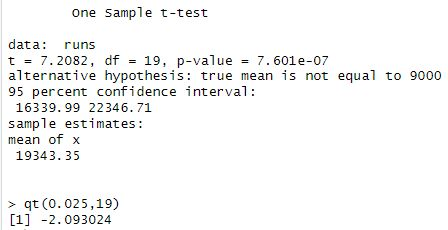
H1: The mean runs scored is not 9000

Rcode:

t.test(runs,alternative="two.sided",mu=9000)

qt(0.025,19)

Output:



Conclusion: Since t(cal)>t(critical), reject null hypothesis at 5% LOS. Thus, the mean runs scored is not 9000.