stat

2022-06-20

This dataset tells details about students name and the marks scored by them in each subjects

```
marks = read.csv("statistics.csv",header = TRUE)
summary(marks)
```

```
##  Serial.number   Student.Name          English         Bengali
##  Min.   : 1.00   Length:20          Min.   :23.00   Min.   :20.00
##  1st Qu.: 5.75   Class :character   1st Qu.:55.25   1st Qu.:54.00
##  Median :10.50   Mode  :character   Median :73.00   Median :70.50
##  Mean   :10.50                      Mean   :70.10   Mean   :65.55
##  3rd Qu.:15.25                      3rd Qu.:86.25   3rd Qu.:80.25
##  Max.   :20.00                      Max.   :96.00   Max.   :91.00
##      Hindi           Maths          History        Geography
##  Min.   :33.00   Min.   :42.00   Min.   :65.00   Min.   : 65.00
##  1st Qu.:43.50   1st Qu.:62.75   1st Qu.:78.50   1st Qu.: 82.75
##  Median :68.50   Median :74.50   Median :86.00   Median : 88.00
##  Mean   :63.85   Mean   :75.75   Mean   :84.60   Mean   : 87.15
##  3rd Qu.:82.00   3rd Qu.:94.25   3rd Qu.:92.25   3rd Qu.: 92.00
##  Max.   :96.00   Max.   :99.00   Max.   :97.00   Max.   :100.00
```

here we select the columns with numerics to perform Statistical Analysis

```
ma = select(marks,c(3:8))
summary(ma)
```

```
##     English         Bengali          Hindi           Maths
##  Min.   :23.00   Min.   :20.00   Min.   :33.00   Min.   :42.00
##  1st Qu.:55.25   1st Qu.:54.00   1st Qu.:43.50   1st Qu.:62.75
##  Median :73.00   Median :70.50   Median :68.50   Median :74.50
##  Mean   :70.10   Mean   :65.55   Mean   :63.85   Mean   :75.75
##  3rd Qu.:86.25   3rd Qu.:80.25   3rd Qu.:82.00   3rd Qu.:94.25
##  Max.   :96.00   Max.   :91.00   Max.   :96.00   Max.   :99.00
##     History        Geography
##  Min.   :65.00   Min.   : 65.00
##  1st Qu.:78.50   1st Qu.: 82.75
##  Median :86.00   Median : 88.00
##  Mean   :84.60   Mean   : 87.15
##  3rd Qu.:92.25   3rd Qu.: 92.00
##  Max.   :97.00   Max.   :100.00
```

```
attach(ma)
```

Hypothesis testing

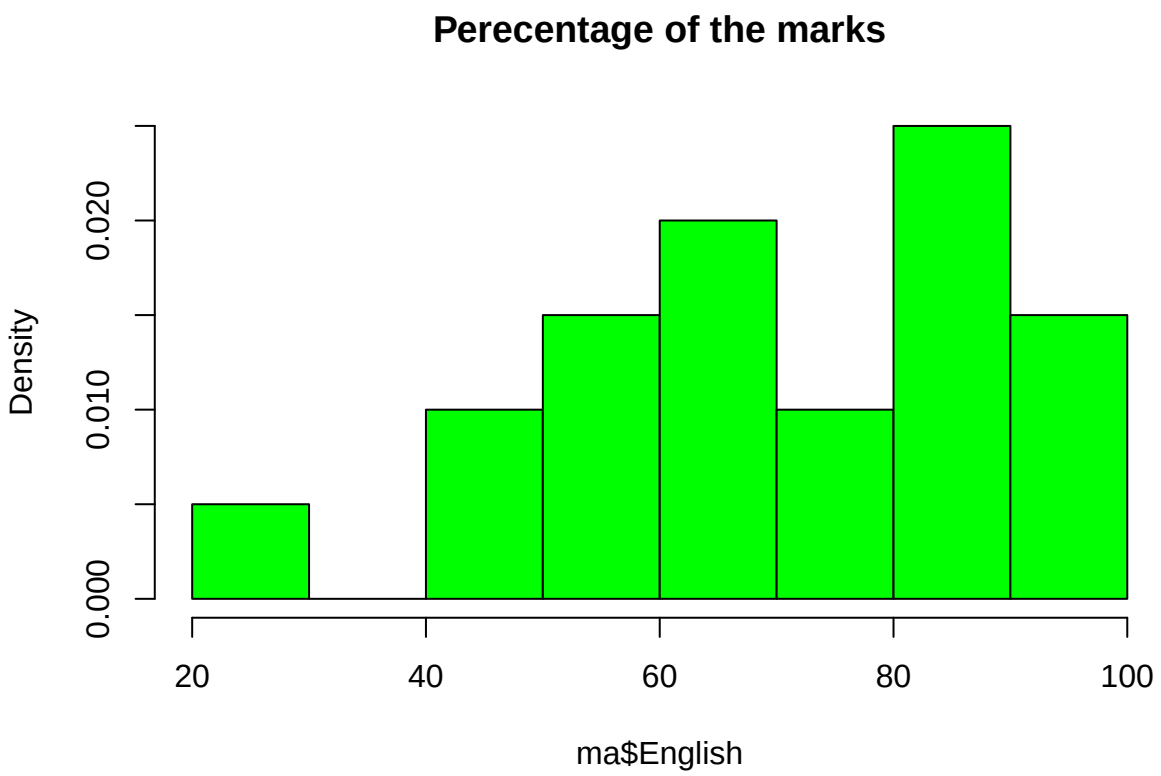
```
mean(marks$English)
```

```
## [1] 70.1
```

```
sd(marks$English)
```

```
## [1] 19.81201
```

```
hist(ma$English,prob=TRUE,col="green",main = "Perecentage of the marks")
```



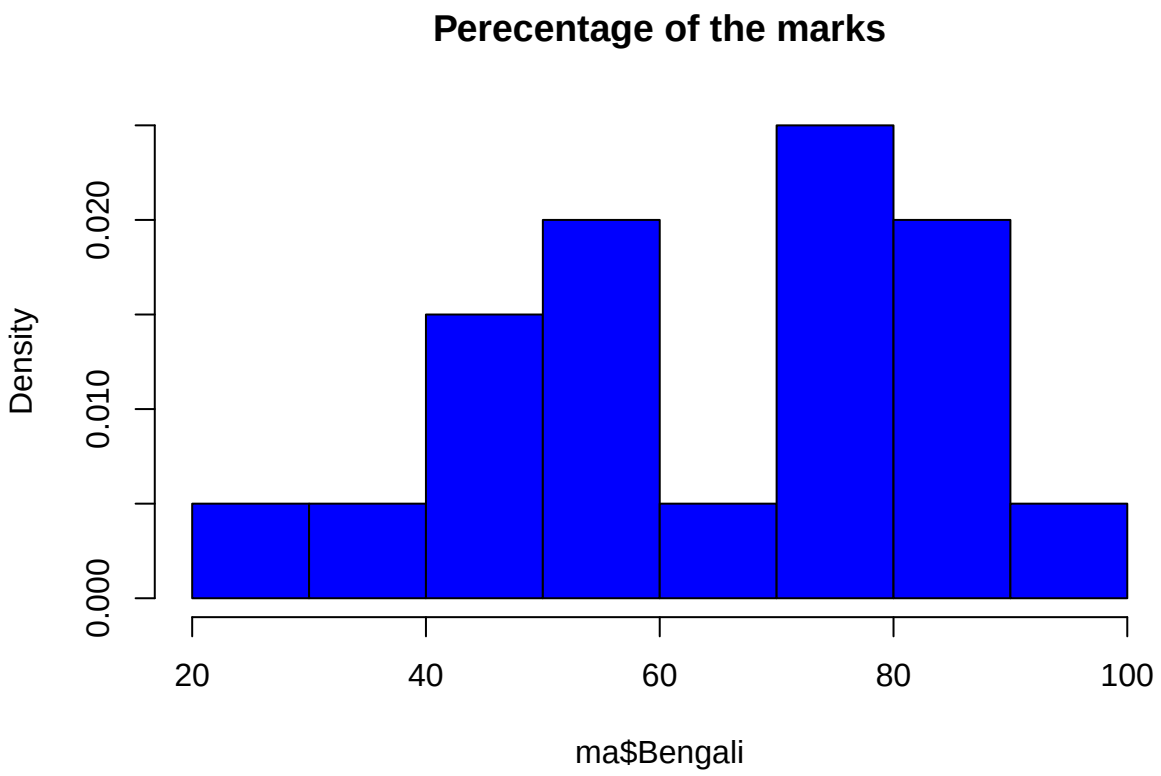

```
mean(marks$Bengali)
```

```
## [1] 65.55
```

```
sd(marks$Bengali)
```

```
## [1] 19.82416
```

```
hist(ma$Bengali,prob=TRUE,col="blue",main = "Perecentage of the marks")
```



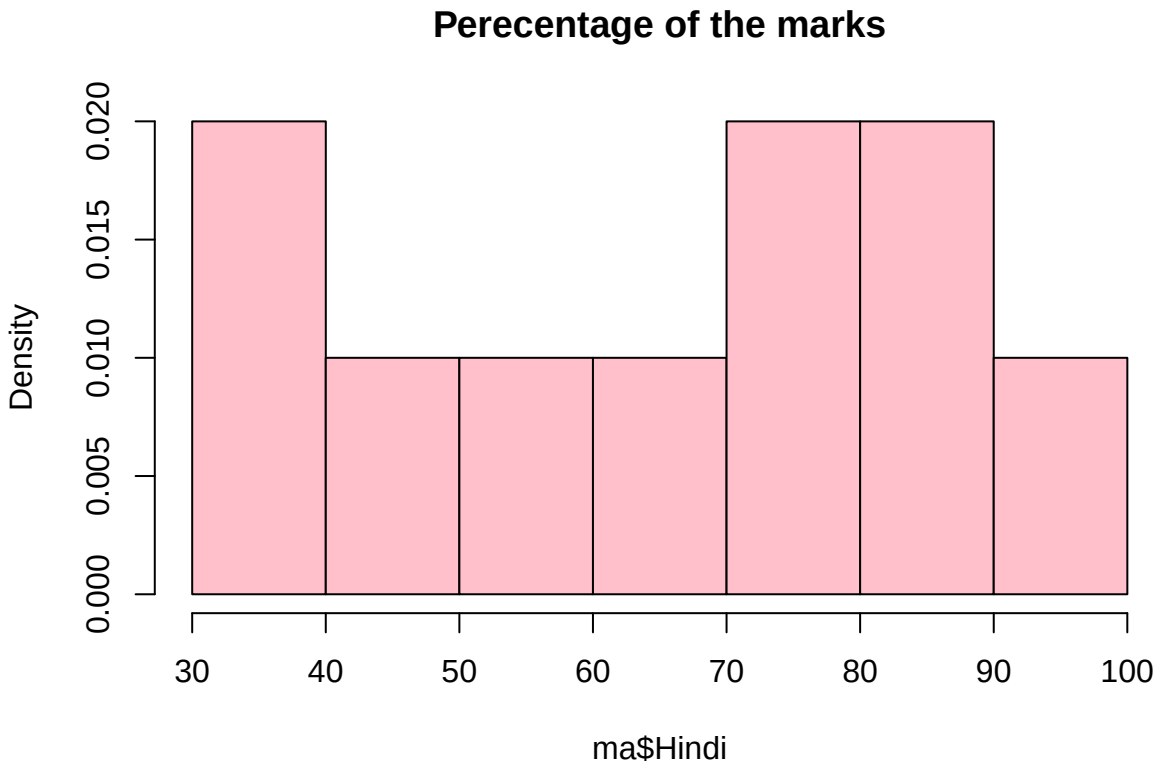

```
mean(marks$Hindi)
```

```
## [1] 63.85
```

```
sd(marks$Hinhi)
```

```
## [1] NA
```

```
hist(ma$Hindi,prob=TRUE,col="pink",main = "Perecentage of the marks")
```



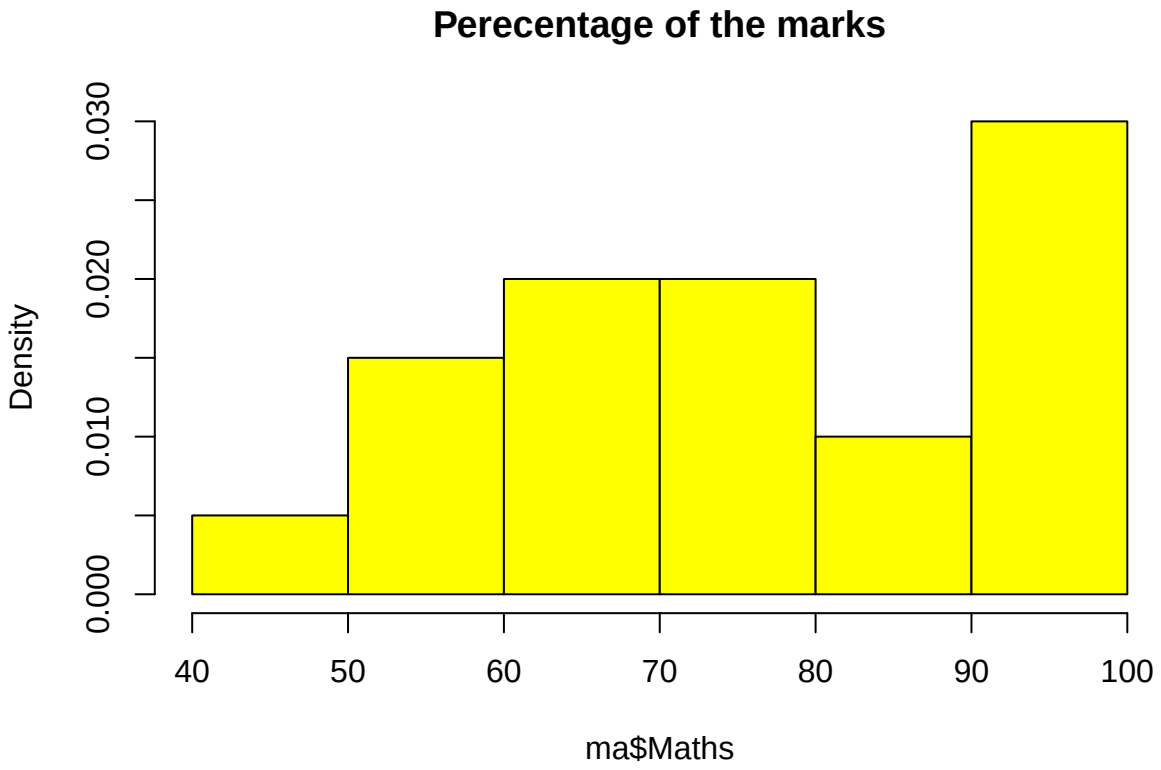

```
mean(marks$Maths)
```

```
## [1] 75.75
```

```
sd(marks$Maths)
```

```
## [1] 17.58551
```

```
hist(ma$Maths,prob=TRUE,col="yellow",main = "Perecentage of the marks")
```



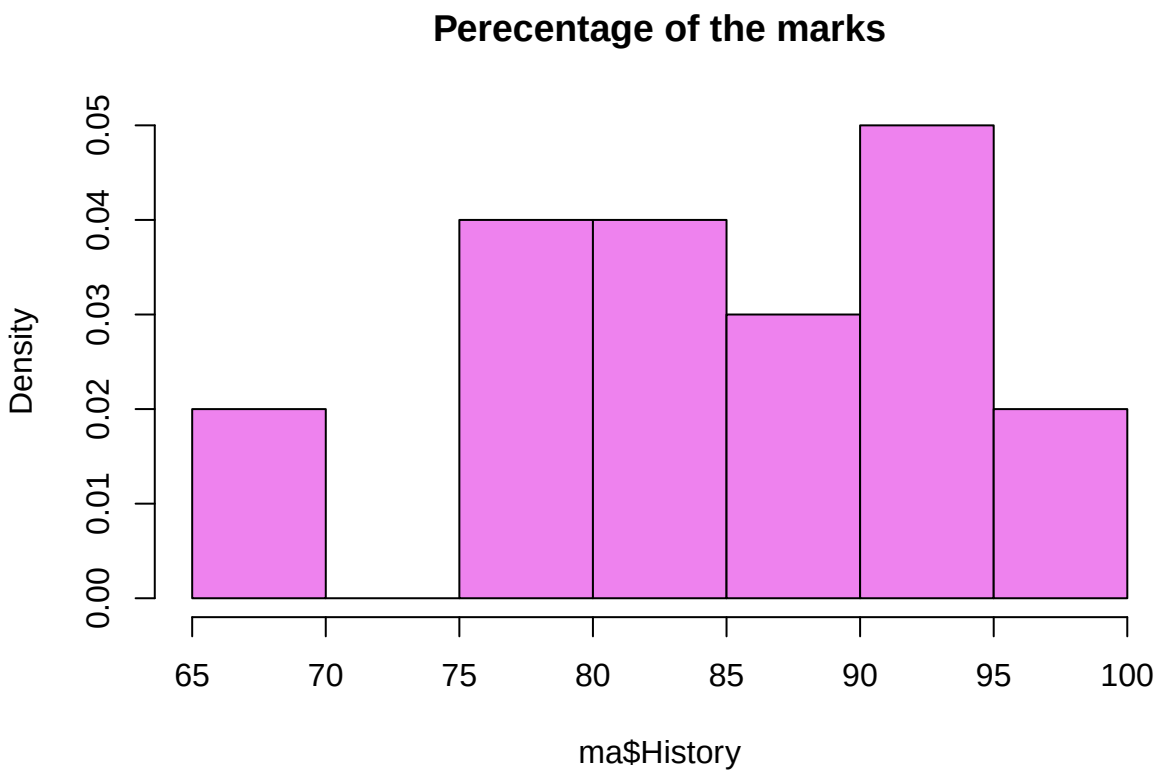

```
mean(marks$History)
```

```
## [1] 84.6
```

```
sd(marks$History)
```

```
## [1] 9.383075
```

```
hist(ma$History,prob=TRUE,col="violet",main = "Perecentage of the marks")
```



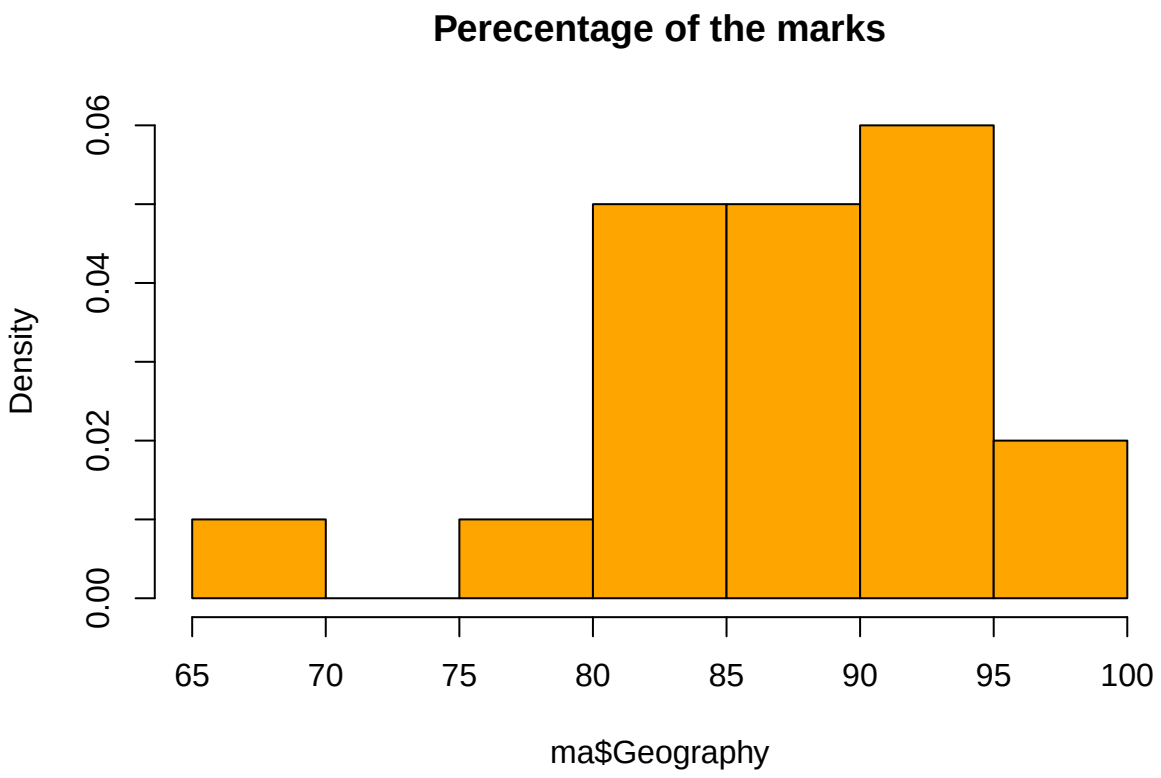

```
mean(marks$Geography)
```

```
## [1] 87.15
```

```
sd(marks$Geography)
```

```
## [1] 8.014953
```

```
hist(ma$Geography,prob=TRUE,col="orange",main = "Perecentage of the marks")
```



Highest percentage of marks scored by students in English is in 80'S. # Highest percentage of the marks scored by students in Bengali is in 75'S. # Highest percentage of the marks scored by students in Hindhi is between 70 to 90. # Highest percentage of the marks scored by students in Maths between 90 to 100 # Highest percentage of the marks scored by students in History is between 90 to 95 # Highest percentage of the marks scored by students in Geography is between 90 to 95

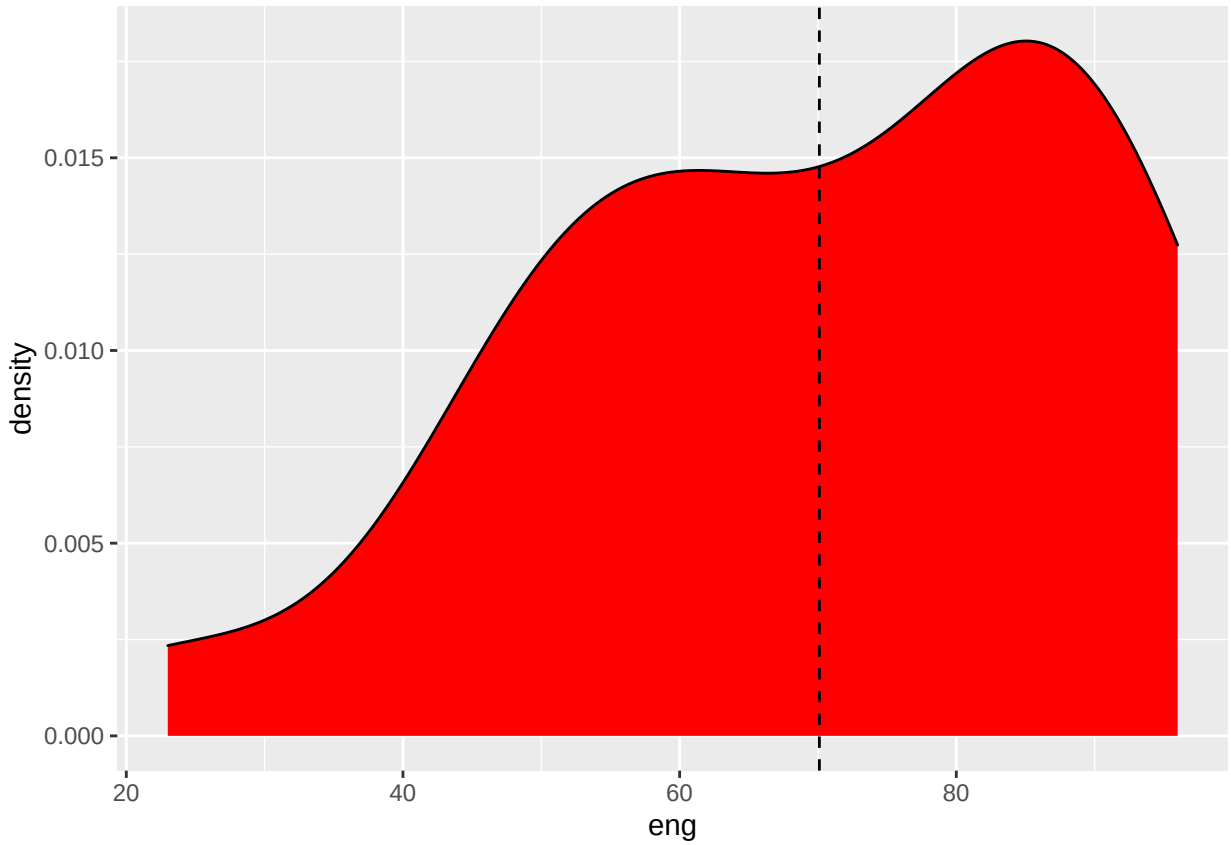
one sample t-test

```
eng = sample(marks$English)
t.test(eng,mu=70.1)
```

```
##
##  One Sample t-test
##
## data:  eng
## t = 0, df = 19, p-value = 1
## alternative hypothesis: true mean is not equal to 70.1
## 95 percent confidence interval:
##  60.82769 79.37231
## sample estimates:
## mean of x
##      70.1
```

```
x1=data.frame(eng)

x1%>%
 ggplot(data=x1,mapping=aes(eng))+geom_density(fill="red")+
  geom_vline(aes(xintercept=mean(eng)),color="black",linetype="dashed",size=0.5)
```

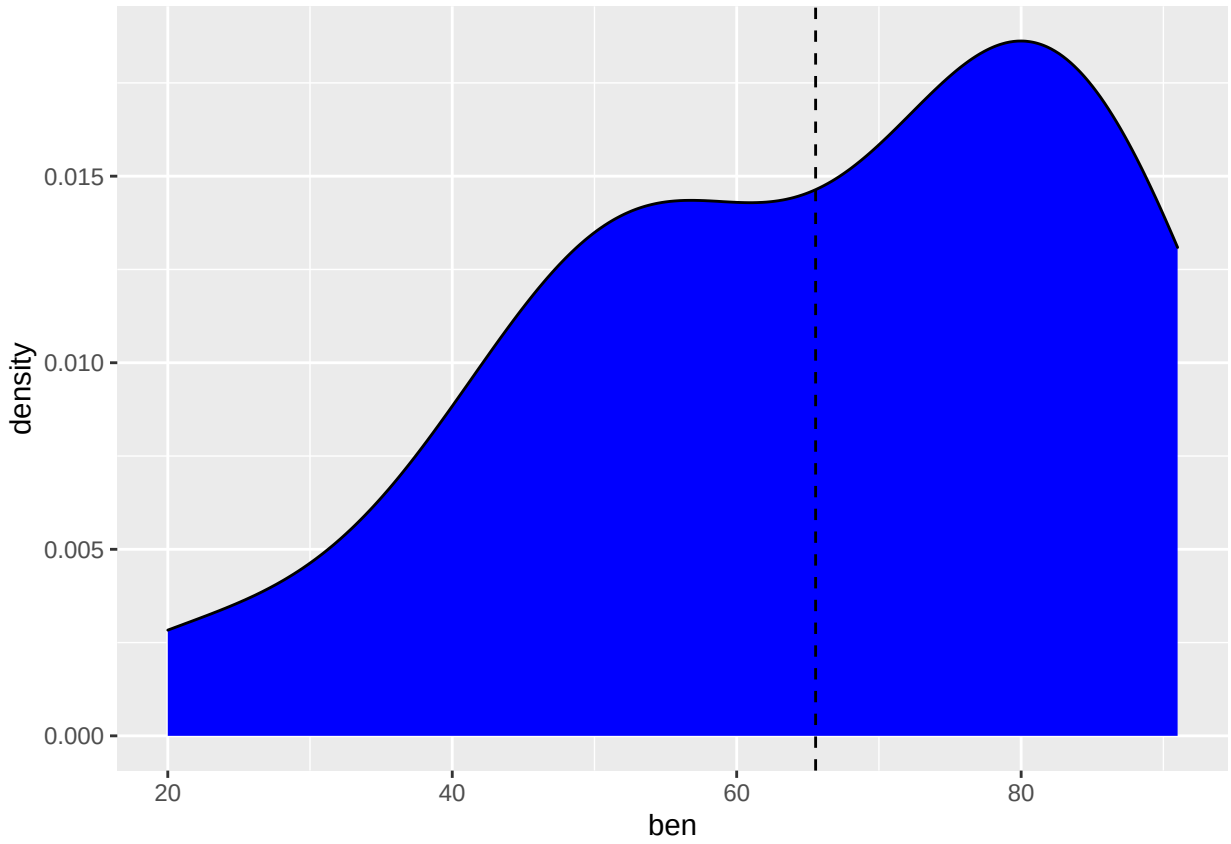



```
ben = sample(marks$Bengali)
t.test(ben,mu=65.55)
```

```
##
##  One Sample t-test
##
## data:  ben
## t = 0, df = 19, p-value = 1
## alternative hypothesis: true mean is not equal to 65.55
## 95 percent confidence interval:
##  56.27201 74.82799
## sample estimates:
## mean of x
##     65.55
```

```
x2=data.frame(ben)

x2%>%
 ggplot(data=x1,mapping=aes(ben))+geom_density(fill="blue")+
  geom_vline(aes(xintercept=mean(ben)),color="black",linetype="dashed",size=0.5)
```

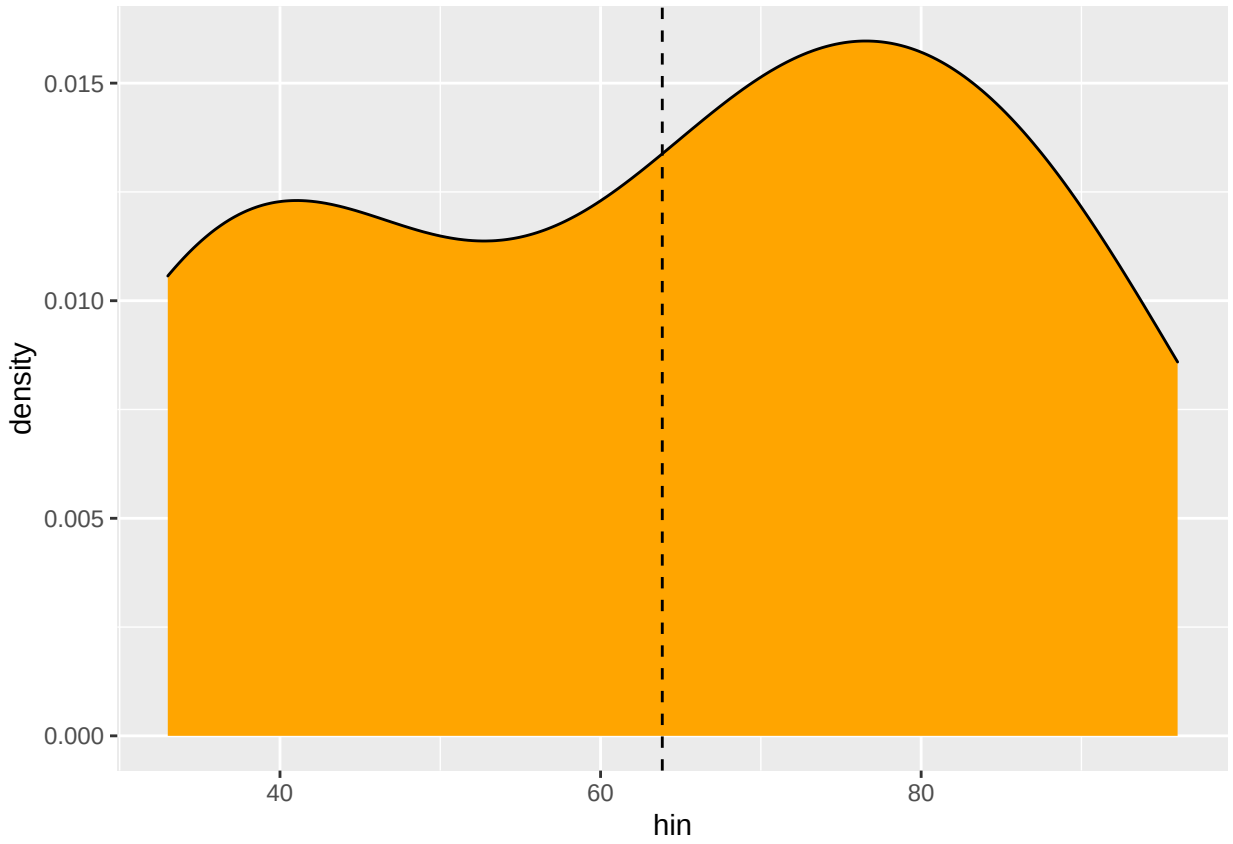



```
hin = sample(marks$Hindi)
t.test(hin,mu=63.85)
```

```
##
##  One Sample t-test
##
## data:  hin
## t = 0, df = 19, p-value = 1
## alternative hypothesis: true mean is not equal to 63.85
## 95 percent confidence interval:
##  53.97399 73.72601
## sample estimates:
## mean of x
##     63.85
```

```
x3=data.frame(hin)

x3%>%
 ggplot(data=x1,mapping=aes(hin))+geom_density(fill="orange")+
  geom_vline(aes(xintercept=mean(hin)),color="black",linetype="dashed",size=0.5)
```

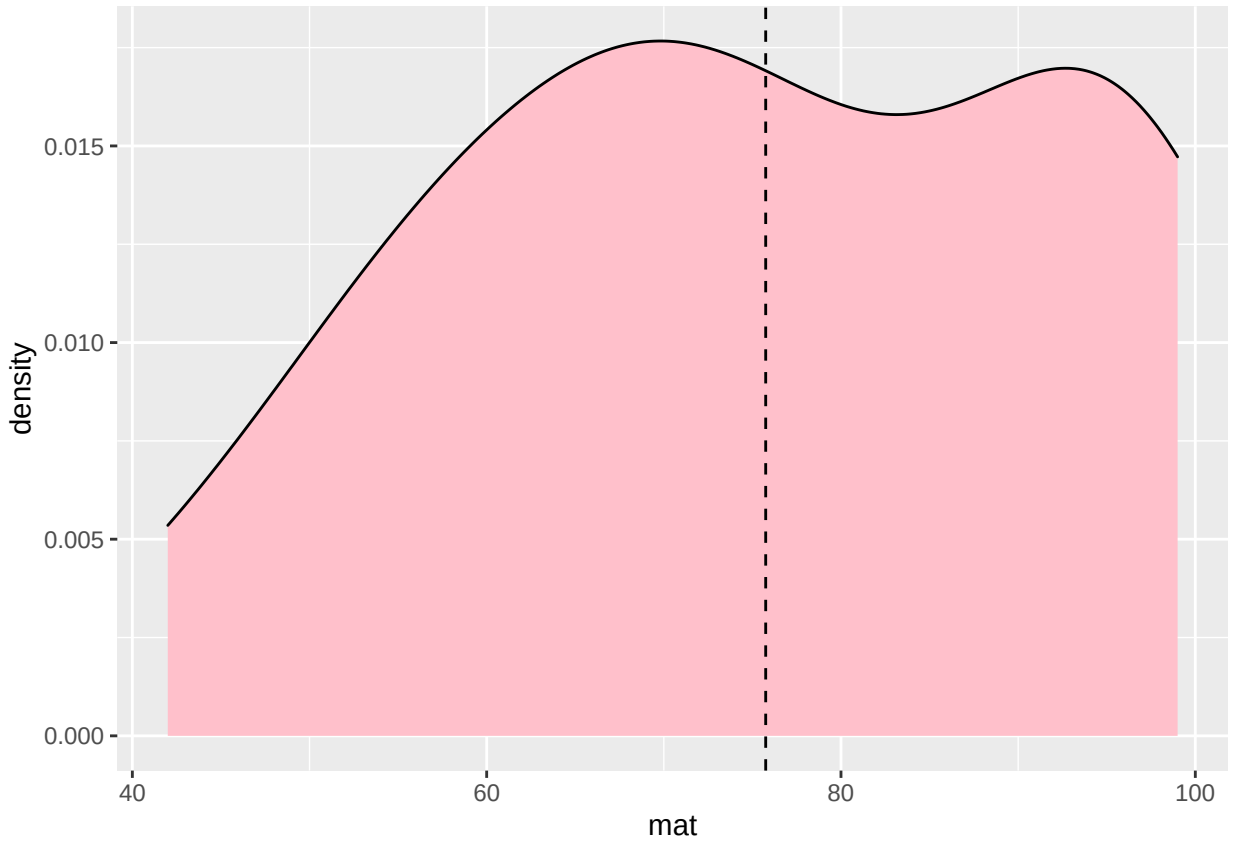



```
mat = sample(marks$Maths)
t.test(mat,mu=75.75)
```

```
##
##  One Sample t-test
##
## data:  mat
## t = 0, df = 19, p-value = 1
## alternative hypothesis: true mean is not equal to 75.75
## 95 percent confidence interval:
##  67.51973 83.98027
## sample estimates:
## mean of x
##     75.75
```

```
x4=data.frame(mat)

x4%>%
 ggplot(data=x1,mapping=aes(mat))+geom_density(fill="pink")+
  geom_vline(aes(xintercept=mean(mat)),color="black",linetype="dashed",size=0.5)
```

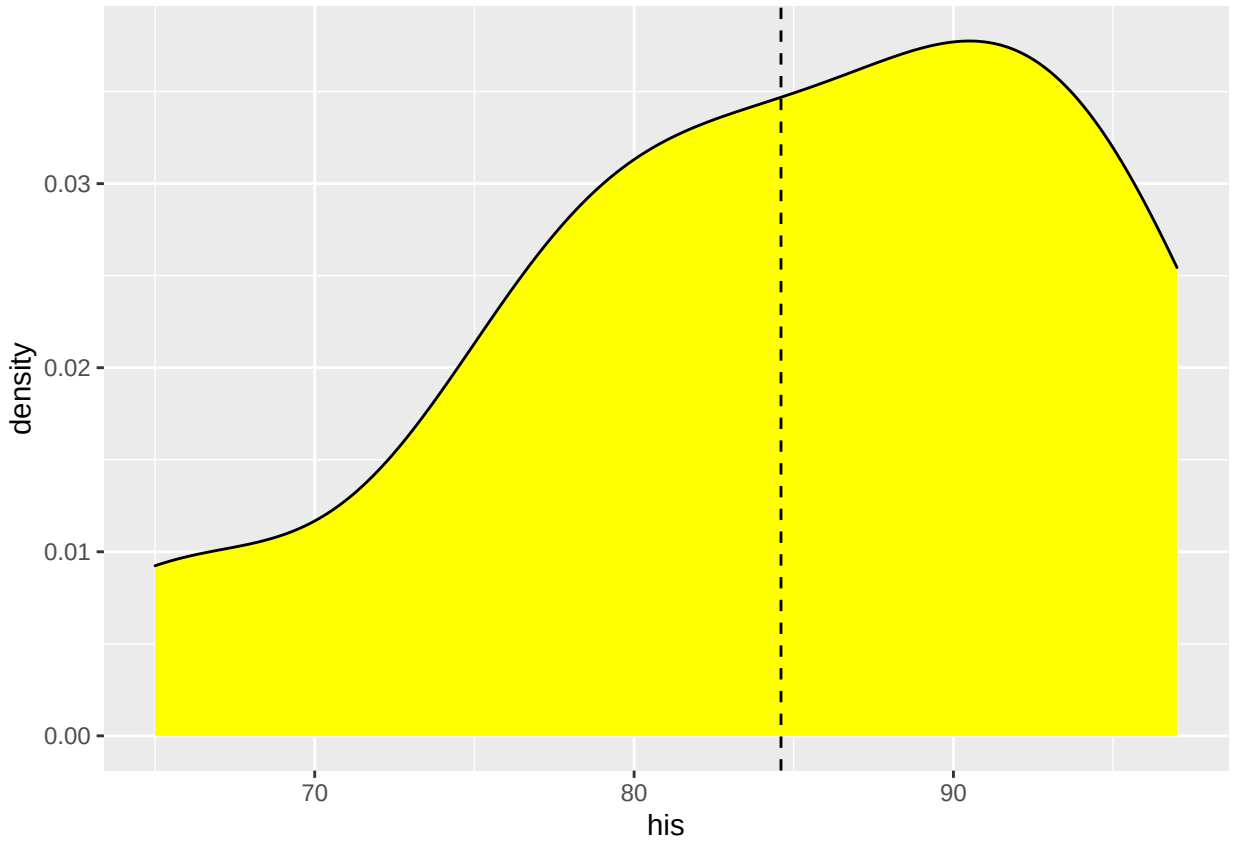



```
his = sample(marks$History)
t.test(his,mu=84.6)
```

```
##
##  One Sample t-test
##
## data:  his
## t = 0, df = 19, p-value = 1
## alternative hypothesis: true mean is not equal to 84.6
## 95 percent confidence interval:
##  80.20859 88.99141
## sample estimates:
## mean of x
##      84.6
```

```
x5=data.frame(his)

x5%>%
 ggplot(data=x1,mapping=aes(his))+geom_density(fill="yellow")+
  geom_vline(aes(xintercept=mean(his)),color="black",linetype="dashed",size=0.5)
```

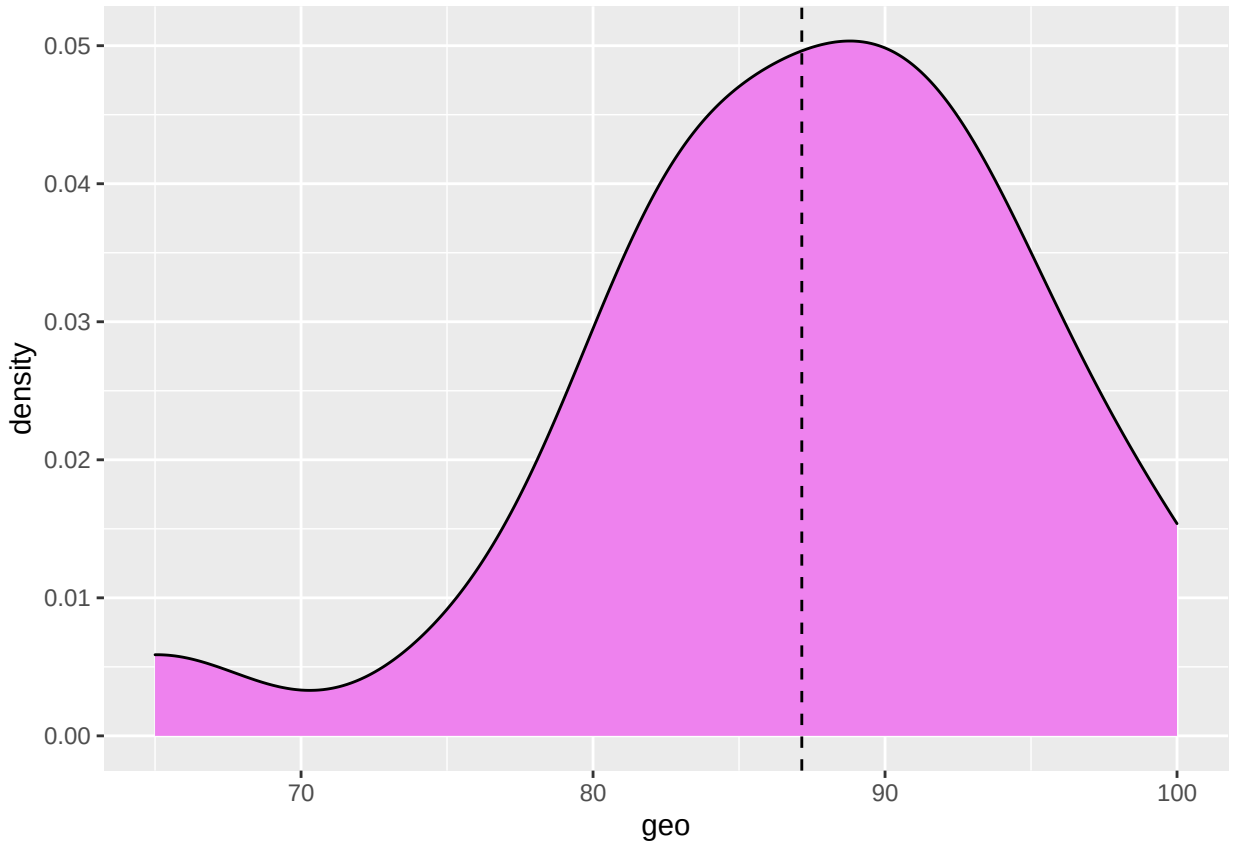



```
geo = sample(marks$Geography)
t.test(geo,mu=87.15)
```

```
##
##  One Sample t-test
##
## data:  geo
## t = 0, df = 19, p-value = 1
## alternative hypothesis: true mean is not equal to 87.15
## 95 percent confidence interval:
##  83.39889 90.90111
## sample estimates:
## mean of x
##     87.15
```

```
x6=data.frame(mat)

x6%>%
 ggplot(data=x1,mapping=aes(geo))+geom_density(fill="violet")+
  geom_vline(aes(xintercept=mean(geo)),color="black",linetype="dashed",size=0.5)
```



From the above result it is observed that p-value for all subject is more than the significant value i.e(1>0.05) the null hypothesis is accepted at 95%,65.55%,53.97%,67.51%,80.20%,83.29% level .

Two sample t-test

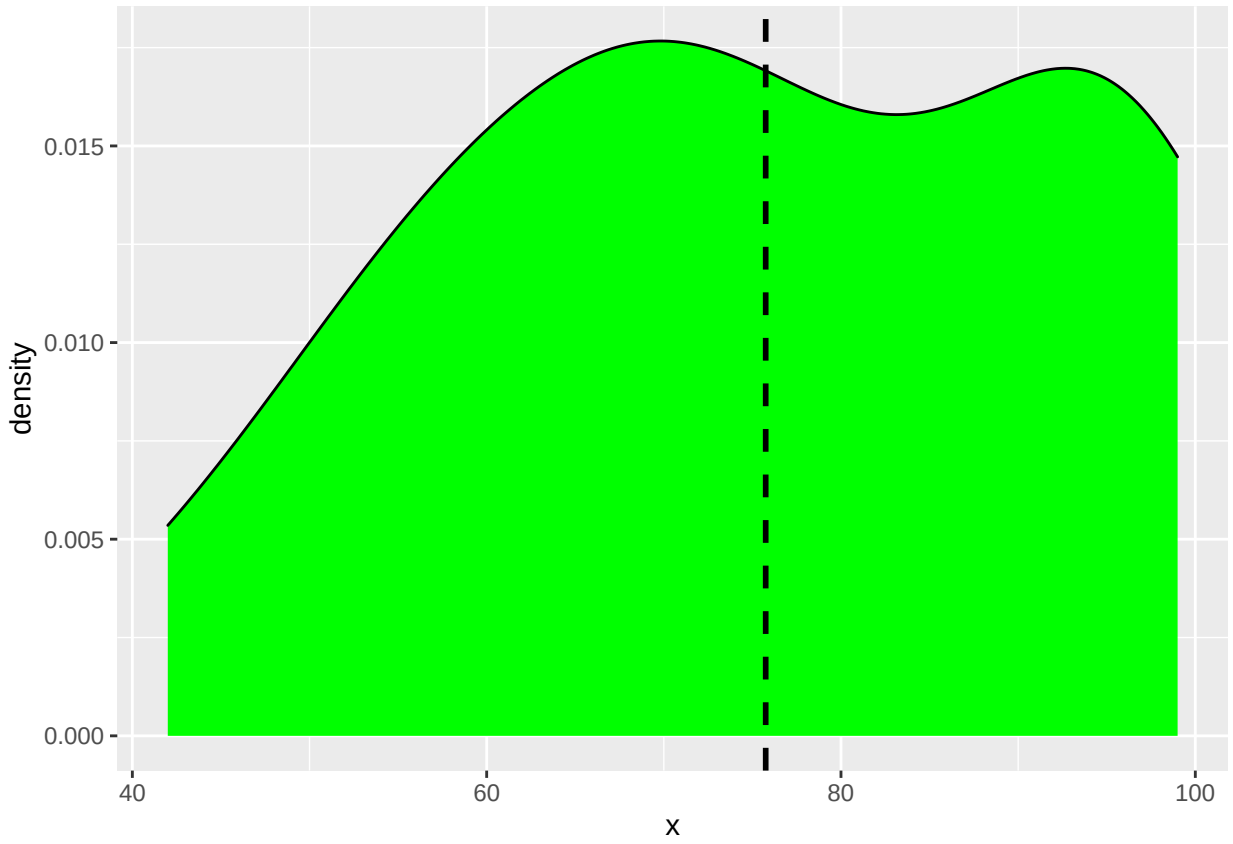
```
x<-sample(ma$Maths)
mean(x)
```

```
## [1] 75.75
```

```
y<-sample(ma$Maths)
mean(y)
```

```
## [1] 75.75
```

```
x1<-data.frame(x)
y1<-data.frame(y)
x1%>%
 ggplot(data=x1,mapping=aes(x))+geom_density(fill="green")+
  geom_vline(aes(xintercept=mean(x)),color="black",linetype="dashed",size=1)
```




```
t.test(x,y,var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  x and y
## t = 0, df = 38, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.25771  11.25771
## sample estimates:
## mean of x mean of y
##     75.75     75.75
```

From the above result it is observed that p-value is more than the significant value i.e(1>0.05) the null hypothesis is accepted at 95% level. Hence there is no significant difference average of two samples.

correlation test

```
cor.test(ma$English,ma$Maths)
```

```
##
##  Pearson's product-moment correlation
##
```

```
## data:  ma$English and ma$Maths
## t = 0.77363, df = 18, p-value = 0.4492
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2858209  0.5761716
## sample estimates:
##       cor
## 0.1793893
```

From the above result it is observed that p-value is greater than the significant value i.e(0.4492>0.05) the null hypothesis is accepted at 5% level.The correlation value is 0.1793893.

INSIGHTS: This dataset tells details about students name and the marks scored by them in each subjects From this data set performing Statistical Analysis we find that # Highest percentage of marks scored by students in English is in 80'S. # Highest percentage of the marks scored by students in Bengali is in 75'S. # Highest percentage of the marks scored by students in Hindhi is between 70 to 90. # Highest percentage of the marks scored by students in Maths between 90 to 100 # Highest percentage of the marks scored by students in History is between 90 to 95 # Highest percentage of the marks scored by students in Geography is between 90 to 95 From the above one sample test result it is observed that p-value for all subject is more than the significant value i.e(1>0.05) the null hypothesis is accepted at 95%,65.55%,53.97%,67.51%,80.20%,83.29% level . From the above two sample test result it is observed that p-value is more than the significant value i.e(1>0.05) the null hypothesis is accepted at 95% level. Hence there is no significant difference average of two samples. From the above correlation test result it is observed that p-value is greater than the significant value i.e(0.4492>0.05) the null hypothesis is accepted at 5% level. The correlation value is 0.1793893