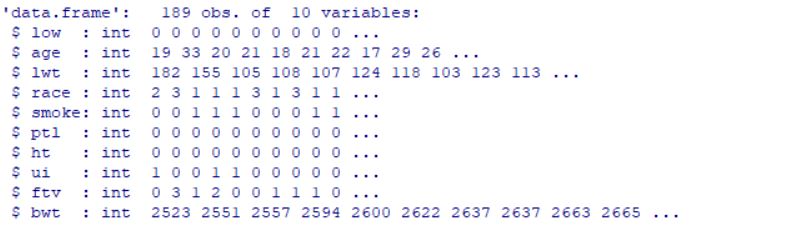
**Take Home Assignment**

1)

library(MASS)

attach(birthwt)



2)

|  |  |  |
| --- | --- | --- |
|  | Scale | Continuous / Discrete |
| age | Interval | Discrete |
| lwt | Ratio | Continuous |
| race | Norminal |  |
| smoke | Norminal |  |
| ptl | Norminal | Discrete |
| ht | Norminal |  |
| ui | Norminal |  |
| ftv | Norminal | Discrete |
| bwt | Ratio | Continuous |
| low | Norminal |  |

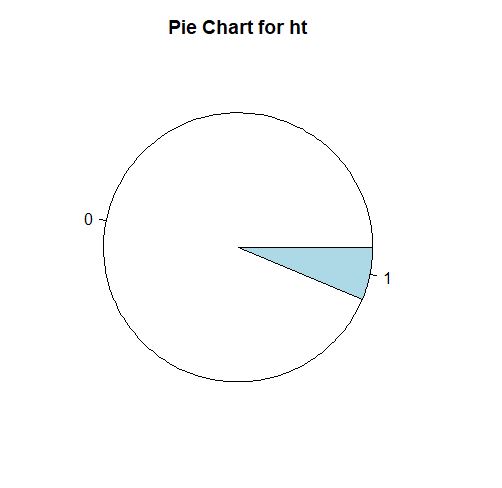
3)

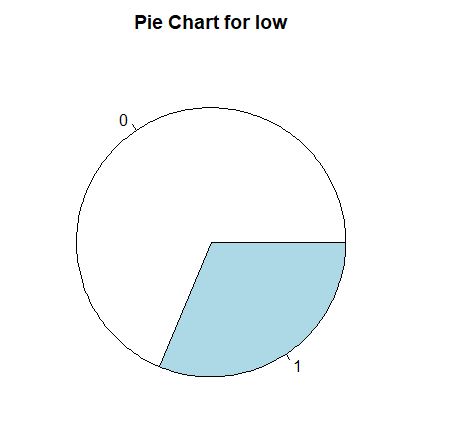
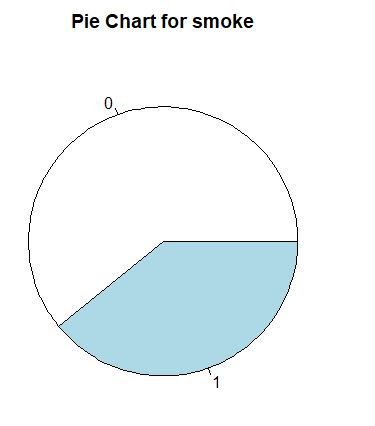
Qualitative variables – ht , smoke , low

pie(table(ht),main = "Pie Chart for ht")

pie(table(smoke),main = "Pie Chart for smoke")

pie(table(low),main = "Pie Chart for low")



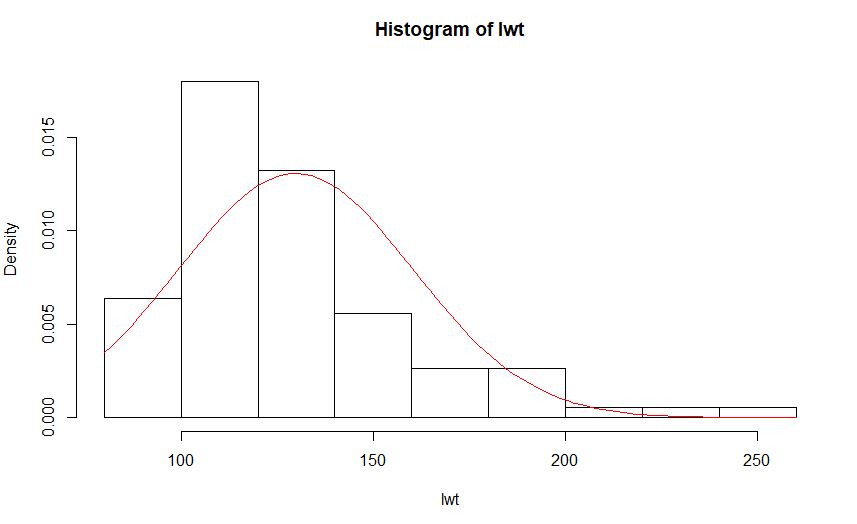


4)

Quantitative variables – age, bwt , lwt, ftv , ptl

hist(lwt,prob=T)

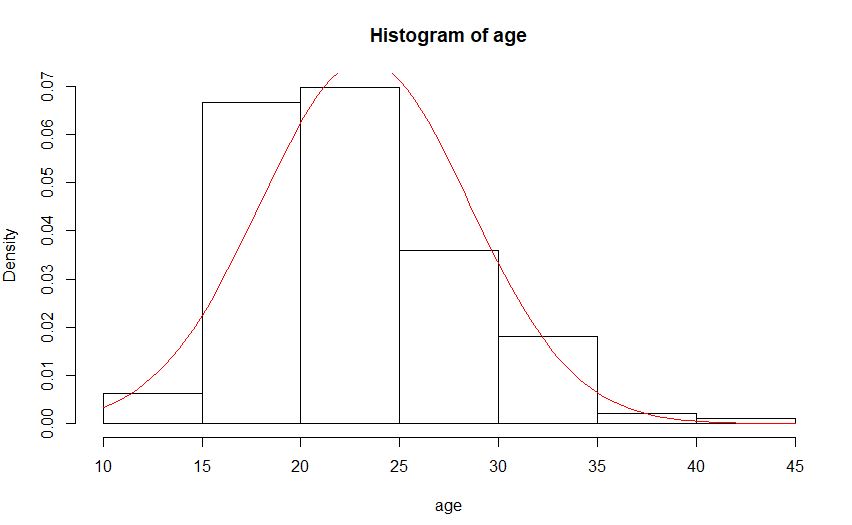
curve(dnorm(x, mean=mean(birthwt$lwt), sd=sd(birthwt$lwt)), col="red", add=TRUE)



* Positive skewness

hist(age,prob=T)

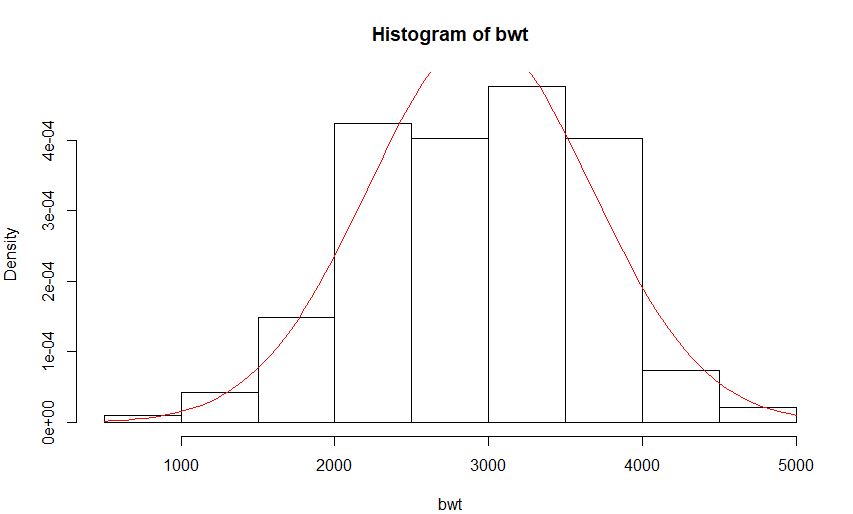
curve(dnorm(x, mean=mean(birthwt$age), sd=sd(birthwt$age)), col="red", add=TRUE)



* Bell Shaped

hist(bwt,prob=T)

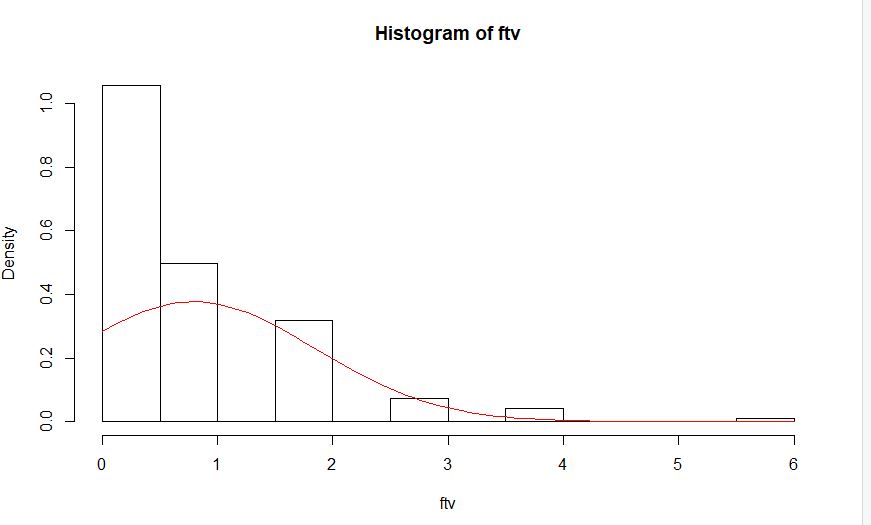
curve(dnorm(x, mean=mean(birthwt$bwt), sd=sd(birthwt$bwt)), col="red", add=TRUE)



* Negative skewness

hist(ftv,prob=T)

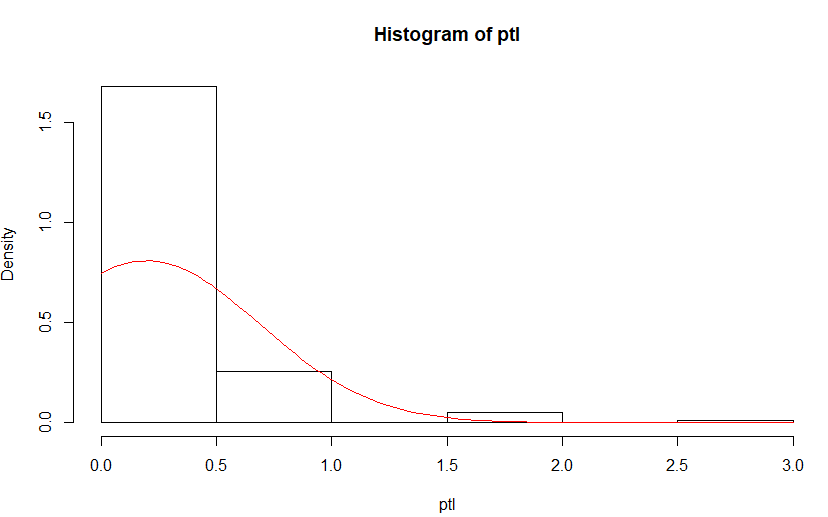
curve(dnorm(x, mean=mean(birthwt$ftv), sd=sd(birthwt$ftv)), col="red", add=TRUE)



* Positive skewness

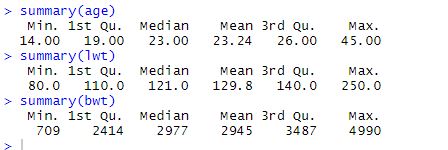
hist(ptl,prob=T)

curve(dnorm(x, mean=mean(birthwt$ptl), sd=sd(birthwt$ptl)), col="red", add=TRUE)



* Positive skewness

5)



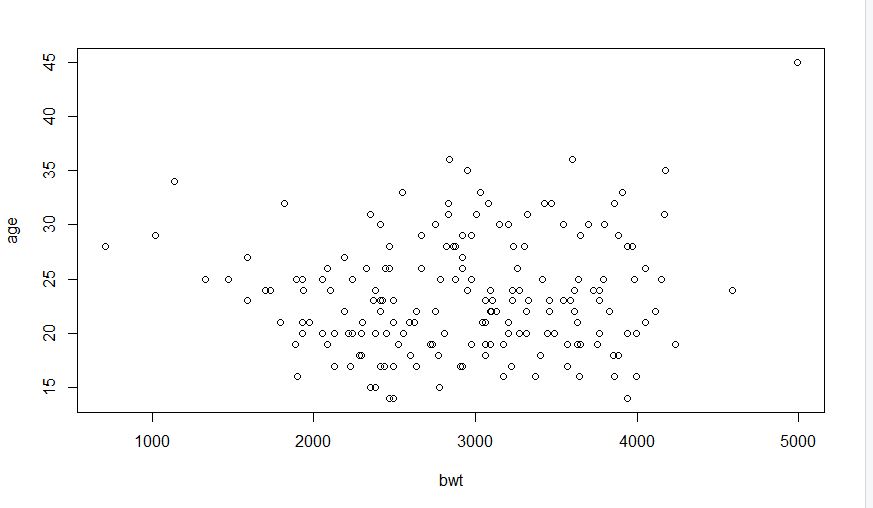
age – mean

bwt – median

lwt – 1st Qu.

6) No relationship

* age and bwt pairs haven’t any linear or non linear pattern.



7)

set.seed(17001031)

8)

birthwt\_df<- data.frame(birthwt)

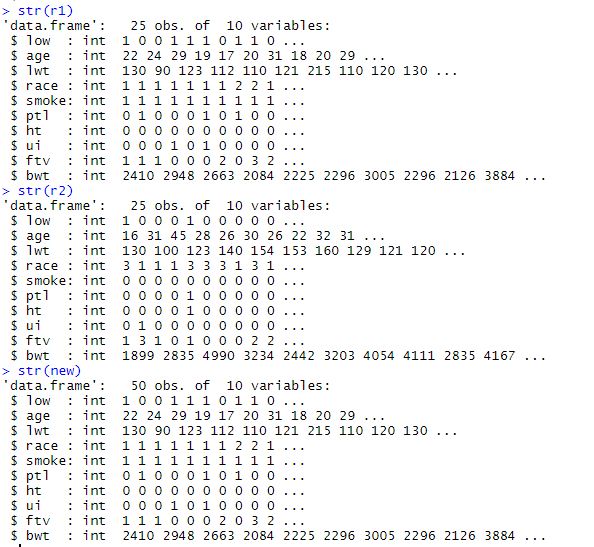
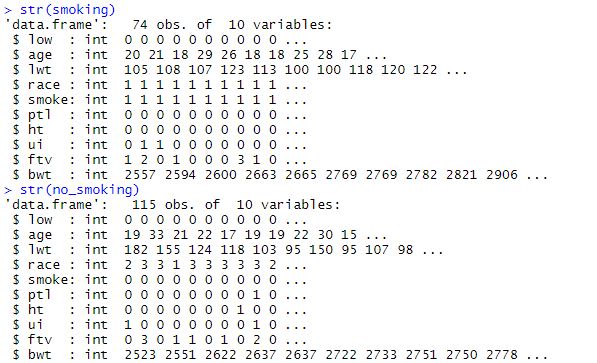
smokeing<- subset.data.frame(birthwt\_df,smoke==1)

no\_smokeing<- subset.data.frame(birthwt\_df,smoke==0)

r1<- smoking[sample(nrow(smoking), 25),]

r2<- no\_smoking[sample(nrow(no\_smoking), 25),]

new<-rbind.data.frame(r1,r2)



9)

new\_m1<-mean(smoking$bwt)

new\_sd1<-sd(smoking$bwt)

new\_m1

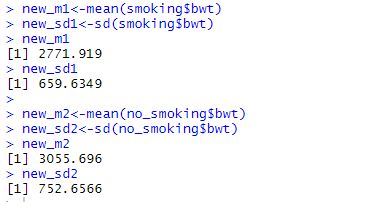
new\_sd1

new\_m2<-mean(no\_smoking$bwt)

new\_sd2<-sd(no\_smoking$bwt)

new\_m2

new\_sd2



10)

𝐻0:𝜇𝑎−𝜇𝑚=0

𝐻𝑎:𝜇𝑎−𝜇𝑚≠0

n1<- nrow(smoking)

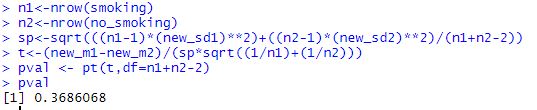
n2<- nrow(no\_smoking)

sp<- sqrt(((n1-1)\*(new\_sd1)\*\*2)+((n2-1)\*(new\_sd2)\*\*2)/(n1+n2-2))

t<- (new\_m1-new\_m2)/(sp\*sqrt((1/n1)+(1/n2))) #finding t value

pval <- pt(t,df=n1+n2-2)

pval



In here pval > 0.05. Hence can not reject the null hypothesis at 5% significant level.

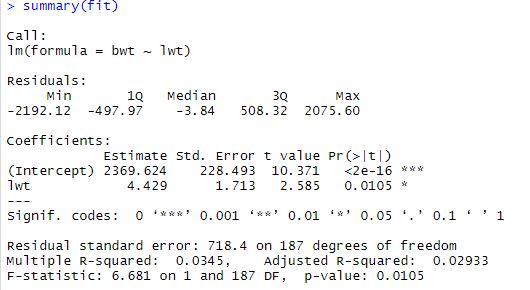
11)

fit<- lm(bwt~lwt)

fit

bwt= 2369.624 + 4.429\*lwt

summary(fit)



𝑟2 is 0.0345, so 3% of the variability in eruptions is explained by the regression fit. It can be considered as a bad fit.