```
library(dplyr)
load(file = "statsample.Rdata")
names(mysample)
dim(df1)
df1 <- mysample[-c(which(mysample$AgeFirstKill == 99999)),] #dropping
missing values recorded as 99999
df1 <- df1[-c(which(is.na(df1$Motive))),]#dropping NA values in Motive
column
df1 <- df1[-c(which((df1$AgeFirstKill + df1$YearBorn) < 1900)),]
df1 <- df1[-c(which(is.na(df1$Sentence))),]#dropping NA values in Motive
column

df1['CareerDuration'] <- (df1$AgeLastKill - df1$AgeFirstKill)
which(df1$CareerDuration < 0) #checking for any unrealistic or negative
values

table(df1$Motive)
table(df1$Race)
table(df1$Sex)
table(df1$Sentence)
table(df1$AgeFirstKill)
mean(df1$AgeFirstKill)
mean(df1$CareerDuration)
sd(df1$AgeFirstKill)
sd(df1$CareerDuration)

###Segregation by motives

anger <- df1[df1$Motive == 'Anger (including mission-oriented killers)',]
revenge <- df1[df1$Motive == 'Revenge or vigilante justice',]
escape <- df1[df1$Motive == 'Escape or avoid arrest',]

dim(anger)
dim(revenge)
dim(escape)


quantile(anger$AgeFirstKill, type = 1)
quantile(revenge$AgeFirstKill, type = 1)
quantile(escape$AgeFirstKill, type = 1)

mean(anger$AgeFirstKill)
mean(revenge$AgeFirstKill)
mean(escape$AgeFirstKill)

median(anger$AgeFirstKill)
median(revenge$AgeFirstKill)
median(escape$AgeFirstKill)


boxplot(anger$AgeFirstKill,
        main = "Motive:Anger (including mission-oriented killers)",
        cex.main=0.9,
        font.main = 4)

boxplot(revenge$AgeFirstKill,
        main = "Motive:Revenge or vigilante justice",
        cex.main=0.9,
        font.main = 4)
```

```
boxplot(escape$AgeFirstKill,
        main = "Motive:Escape or avoid arrest",
        cex.main=0.9,
        font.main = 4)


#Modelling distributions and density curves
par(mfrow = c(1, 2))

 hist(anger$AgeFirstKill ,
     xlab = "Age at First kill",
     main = "Histogram of Motive:Anger (including mission-oriented
killers)",
     cex.main=0.9,
     font.main = 4,
     freq = FALSE,
     xlim = (c(0,70))
     )
mu <- mean(anger$AgeFirstKill)
sigma <- sd(anger$AgeFirstKill)
x <- seq(from = 0, to = 70, by = 0.1)
lines(x, dnorm(x, mean = mu, sd = sigma), lwd = 2, col = "blue")


#Modelling distributions and density curves revenge
hist(revenge$AgeFirstKill ,
     xlab = "Age at First kill",
     main = "Histogram of Motive:Revenge or vigilante justice)",
     cex.main=0.9,
     font.main = 4,
     freq = FALSE,
     xlim = (c(0,70))
     )

mu <- mean(revenge$AgeFirstKill)
sigma <- sd(revenge$AgeFirstKill)
x <- seq(from = 0, to = 65, by = 0.1)
lines(x, dnorm(x, mean = mu, sd = sigma), lwd = 2, col = "blue")



#Modelling distributions and density curves Escape

hist(escape$AgeFirstKill ,
     xlab = "Age at First kill",
     main = "Histogram plot of Motive:Escape or avoid arrest)",
     cex.main=0.9,
     font.main = 4,
     freq = FALSE,
     xlim = (c(10,65))
)

mu <- mean(escape$AgeFirstKill)
sigma <- sd(escape$AgeFirstKill)
x <- seq(from = 15, to = 60, by = 0.1)
lines(x, dnorm(x, mean = mu, sd = sigma), lwd = 2, col = "blue")

#Suitability of model checking using CDF and ks test
#CDF
Fn <- ecdf(anger$AgeFirstKill)
```

```
mu <- mean(anger$AgeFirstKill)
sigma <- sd(anger$AgeFirstKill)
G <- function(x){return(pnorm(x, mean = mu, sd = sigma))}
plot(Fn, verticals = TRUE, pch = NA, xlab="Age at First kill",
     ylab = "Fn (Age at First kill)",
     main = "CDF of Motive:Anger (including mission-oriented killers)",
     cex.main=0.9,
     font.main = 4,)
x <- 1:100
lines(x, G(x), col = "red3")

# Perform the Kolmogorov-Smirnov test:

ks.test(x =  anger$AgeFirstKill,
        y = "pnorm",
        mean = mu, sd = sigma)

#CDF for motive revenge

Fn <- ecdf(revenge$AgeFirstKill)
mu <- mean(revenge$AgeFirstKill)
sigma <- sd(revenge$AgeFirstKill)
G <- function(x){return(pnorm(x, mean = mu, sd = sigma))}
plot(Fn, verticals = TRUE, pch = NA,xlab="Age at First kill",
     ylab = "Fn (Age at First kill)",
     main = "CDF of Motive:Revenge or vigilante justice",
     cex.main=0.9,
     font.main = 4)
x <- 1:100
lines(x, G(x), col = "red3")

# Perform the Kolmogorov-Smirnov test: for revenge

ks.test(x =  revenge$AgeFirstKill,
        y = "pnorm",
        mean = mu, sd = sigma)

#CDF for escape motive
Fn <- ecdf(escape$AgeFirstKill)
mu <- mean(escape$AgeFirstKill)
sigma <- sd(escape$AgeFirstKill)
G <- function(x){return(pnorm(x, mean = mu, sd = sigma))}
plot(Fn, verticals = TRUE, pch = NA,xlab="Age at First kill",
     ylab = "Fn (Age at First kill)",
     main = "CDF of Motive:Escape or avoid arrest",
     cex.main=0.9,
     font.main = 4)
x <- 1:100
lines(x, G(x), col = "red3")

# Perform the Kolmogorov-Smirnov test:for escape motive

ks.test(x =  escape$AgeFirstKill,
        y = "pnorm",
        mean = mu, sd = sigma)


# Normality check using Q-Q plots and shapiro-wilks test
#motive anger
mu <- mean(anger$AgeFirstKill)
```

```
sigma <- sd(anger$AgeFirstKill)
qqnorm(anger$AgeFirstKill,
       main = "Q-Q plot of Motive:Anger (including mission-oriented
killers)"
               ,cex.main=0.9)
abline(a = mu, b = sigma, col = "red")
shapiro.test(anger$AgeFirstKill)

#motive Revenge
mu <- mean(revenge$AgeFirstKill)
sigma <- sd(revenge$AgeFirstKill)
qqnorm(revenge$AgeFirstKill,main = "Q-Q plot of Motive:Revenge or vigilante
justice")
abline(a = mu, b = sigma, col = "red")
shapiro.test(revenge$AgeFirstKill)


#motive escape
mu <- mean(escape$AgeFirstKill)
sigma <- sd(escape$AgeFirstKill)
qqnorm(escape$AgeFirstKill,main = "Q-Q plot of Motive:Escape or avoid
arrest")
abline(a = mu, b = sigma, col = "red")
shapiro.test(escape$AgeFirstKill)



# Estimating mu and sigma for Normal Distribution motive Anger

muhat1 <- mean(anger$AgeFirstKill)
muhat2 <- quantile(anger$AgeFirstKill, type = 1)[3]
par(mfrow = c(1, 2))

hist(anger$AgeFirstKill, freq = FALSE ,xlab="Age at First kill",
     main = "Anger (including mission-oriented killers)",
     cex.main=0.9,
     font.main = 4)
abline(v = muhat1, col = "red3", lwd = 3)
abline(v = 27, col = "blue", lty = 2, lwd = 3)

hist(anger$AgeFirstKill, freq = FALSE,xlab="Age at First kill",
     main = "Anger (including mission-oriented killers)",
     cex.main=0.9,
     font.main = 4)
abline(v = muhat1, col = "red3", lwd = 3)
abline(v = muhat2, col = "blue", lty = 2, lwd = 3)

#Estimating sigma for motive Anger
sigma2hat1 <- sd(anger$AgeFirstKill)^2
sigma2hat2 <- (186/187)*sd(anger$AgeFirstKill)^2
par(mfrow = c(1, 2))

hist(anger$AgeFirstKill ,freq = FALSE, xlim = c(0,100),xlab="Age at First
kill",
     main = "Anger (including mission-oriented killers)",
     cex.main=0.9,
     font.main = 4)
abline(v = sigma2hat1, col = "red3", lwd = 3)
abline(v = 74, col = "blue", lty = 2, lwd = 3)
```

```
hist(anger$AgeFirstKill ,freq = FALSE,xlim = c(0,100),,xlab="Age at First
kill",
     main = "Anger (including mission-oriented killers)",
     cex.main=0.9,
     font.main = 4)
abline(v = sigma2hat1, col = "red3", lwd = 3)
abline(v = sigma2hat2, col = "blue", lty = 2, lwd = 3)



# Estimating mu and sigma for Normal Distribution motive revenge

muhat1 <- mean(revenge$AgeFirstKill)
muhat2 <- quantile(revenge$AgeFirstKill, type = 1)[3]
par(mfrow = c(1, 2))

hist(revenge$AgeFirstKill, freq = FALSE,xlab="Age at First kill",
     ylab = "(Age at First kill)",
     main = "Revenge or vigilante justice",
     cex.main=0.9,
     font.main = 4)
abline(v = muhat1, col = "red3", lwd = 3)
abline(v = 27, col = "blue", lty = 2, lwd = 3)

hist(revenge$AgeFirstKill, freq = FALSE,xlab="Age at First kill",
     ylab = "(Age at First kill)",
     main = "Revenge or vigilante justice",
     cex.main=0.9,
     font.main = 4)
abline(v = muhat1, col = "red3", lwd = 3)
abline(v = muhat2, col = "blue", lty = 2, lwd = 3)

#Estimating sigma for motive revenge
sigma2hat1 <- sd(revenge$AgeFirstKill)^2
sigma2hat2 <- (57/58)*sd(revenge$AgeFirstKill)^2

par(mfrow = c(1, 2))

hist(revenge$AgeFirstKill ,freq = FALSE, xlim = c(0,100),xlab="Age at First
kill",
     main = "Revenge or vigilante justice",
     cex.main=0.9,
     font.main = 4)
abline(v = sigma2hat1, col = "red3", lwd = 3)
abline(v = 74, col = "blue", lty = 2, lwd = 3)

hist(revenge$AgeFirstKill ,freq = FALSE,xlim = c(0,100),xlab="Age at First
kill",
     main = "Revenge or vigilante justice",
     cex.main=0.9,
     font.main = 4)
abline(v = sigma2hat1, col = "red3", lwd = 3)
abline(v = sigma2hat2, col = "blue", lty = 2, lwd = 3)


# Estimating mu and sigma for Normal Distribution motive escape

muhat1 <- mean(escape$AgeFirstKill)
muhat2 <- quantile(escape$AgeFirstKill, type = 1)[3]
par(mfrow = c(1, 2))
```

```
hist(escape$AgeFirstKill, freq = FALSE,xlab="Age at First kill",
     main = "Revenge or vigilante justice",
     cex.main=0.9,
     font.main = 4)
abline(v = muhat1, col = "red3", lwd = 3)
abline(v = 27, col = "blue", lty = 2, lwd = 3)

hist(escape$AgeFirstKill, freq = FALSE,xlab="Age at First kill",
     main = "Revenge or vigilante justice",
     cex.main=0.9,
     font.main = 4)
abline(v = muhat1, col = "red3", lwd = 3)
abline(v = muhat2, col = "blue", lty = 2, lwd = 3)

#Estimating sigma for motive escape
sigma2hat1 <- sd(escape$AgeFirstKill)^2
sigma2hat2 <- (20/21)*sd(escape$AgeFirstKill)^2

par(mfrow = c(1, 2))

hist(escape$AgeFirstKill ,freq = FALSE, xlim = c(0,80),xlab="Age at First
kill",
     main = "Escape or avoid arrest",
     cex.main=0.9,
     font.main = 4)
abline(v = sigma2hat1, col = "red3", lwd = 3)
abline(v = 74, col = "blue", lty = 2, lwd = 3)

hist(escape$AgeFirstKill ,freq = FALSE,xlim = c(0,80),xlab="Age at First
kill",
     main = "Escape or avoid arrest",
     cex.main=0.9,
     font.main = 4)
abline(v = sigma2hat1, col = "red3", lwd = 3)
abline(v = sigma2hat2, col = "blue", lty = 2, lwd = 3)


#Inferences and Test Statistic
len1<- length(anger$AgeFirstKill)
len2<- length(revenge$AgeFirstKill)
len3<- length(escape$AgeFirstKill)

varience <- 74 #known value from past research
x <-mean(anger$AgeFirstKill)
y <-mean(revenge$AgeFirstKill)
v <-mean(escape$AgeFirstKill)
x
y
v

#taking mu as 27
z = (x-27)/sqrt(varience/len1)
p <- 2 * pnorm(z)
p
# alpha value 5%
c <- qnorm(0.975)
(z >  c) | (z < -c)

# motive revenge
```

```
z = (y-27)/sqrt(varience/len2)
p <- 2 * pnorm(z)
p
# alpha value 5%
c <- qnorm(0.975)
(z >  c) | (z < -c)

# motive escape
z = (v-27)/sqrt(varience/len3)
p <- 2 * pnorm(z)
p
# alpha value 5%
c <- qnorm(0.975)
(z >  c) | (z < -c)

#confidence intervals
CI1 =  x + c(-1, 1)*1.96*sqrt(varience/len1)
CI2 =  y + c(-1, 1)*1.96*sqrt(varience/len2)
CI3 =  v + c(-1, 1)*1.96*sqrt(varience/len3)
CI1
CI2
CI3
```