TRAFFIC<-read.csv('TRAFFIC.csv')

lincoln.data <- subset(TRAFFIC, TRAFFIC$TUNNEL == "Lincoln")

nrow(lincoln.data)

holland.data <- subset(TRAFFIC, TRAFFIC$TUNNEL == "Holland")

#traffic at lincoln

lincoln.traffic <- lincoln.data$VOLUME\_PER\_MINUTE

length(lincoln.traffic)

#traffic at holland

holland.traffic <- holland.data$VOLUME\_PER\_MINUTE

# standard deviation of two samples

sd.lincoln <- sd(lincoln.traffic)

sd.holland <- sd(holland.traffic)

# means of two samples

mean.lincoln <- mean(lincoln.traffic)

mean.lincoln

mean.holland <- mean(holland.traffic)

mean.holland

#length of lincoln and holland

len\_lincoln <- length(lincoln.traffic)

len\_holland <- length(holland.traffic)

#standard deviation of difference population

sd.lin.hol <- sqrt(sd.lincoln^2/len\_lincoln + sd.holland^2/len\_holland)

sd.lin.hol

#z score

zeta <- (mean.lincoln - mean.holland)/sd.lin.hol

#plot red line

plot(x=seq(from = -5, to= 5, by=0.1),y=dnorm(seq(from = -5, to= 5, by=0.1),mean=0),type='l',xlab = 'mean difference', ylab='possibility')

abline(v=zeta, col='red')

#get p

p = 1-pnorm(zeta)

p