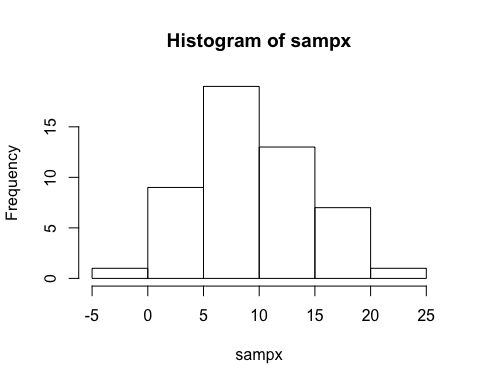
Bootstrap

Clinton Tippett

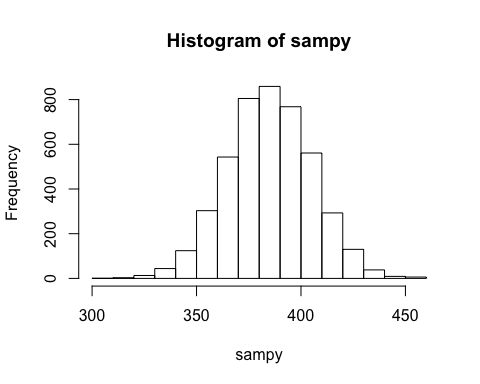
June 6, 2016

# I am going to create four vectors, two which will have a normal distribution and two an exponential distribution. I will then use these vectors to perform the bootstrap technique in order to show examples of Central Limit Theorem.

sampx <- rnorm(50,10,5)  
sampy <- rnorm(4500,385,20)  
hist(sampx)



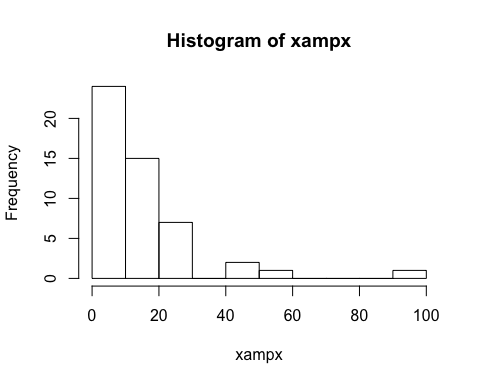
hist(sampy)



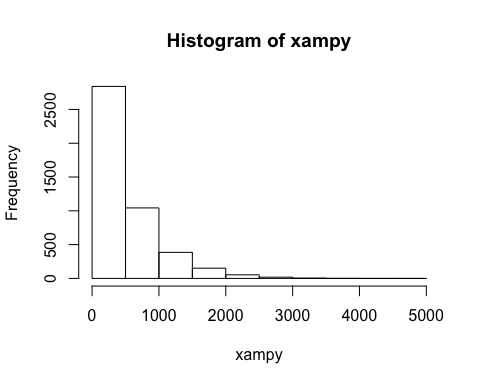
# I created two vectors with a normal distribution. Each of these are unique, and contain vastly different sizes and amounts.

# 'sampx' and 'sampy' are the names of the two vectors.

xampx <- rexp(50, 1/20)  
hist(xampx)



xampy <- rexp(4500, 1/500)  
hist(xampy)



# Here I have created two new vectors 'xampx' and 'xampy'.

# These two vectors are of an exponential distribution

# I used to 'rexp' function to generate the numbers with a defined mean.

# I then tested the vectors using a histogram to verify that the distribution was indeed exponential.

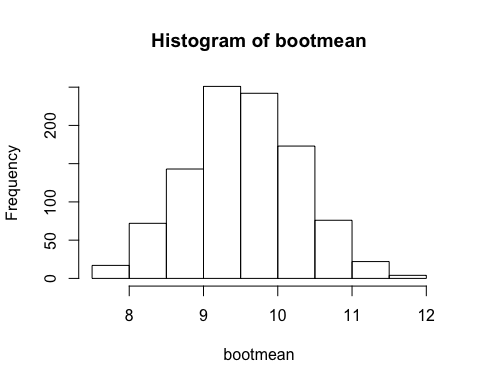
# Now that we have four vectors, two of normal and two of exponential dist, we can apply bootstap technique to verify that the outcomes will follow central limit theorem

# Next, I am going to use the loop function to produce several automated iterations and calculate the means of each iterations

R <- 1000  
bootmean <- numeric(R)  
for (i in 1:R) {  
bootsample <- sample(sampx, size=length(sampx), replace=TRUE)  
bootmean[i] <- mean(bootsample)}

# now, I am going to check the normality of the result 'bootmean'

hist(bootmean)

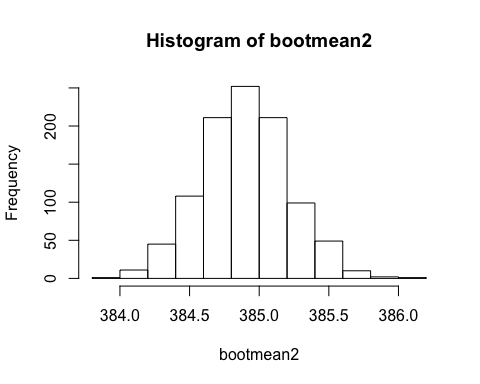


# as expected, the results from bootmean for a normal distribution.

# this was doubly expected -- The original vector from which the sample was obtained was a normal distribution

# Now I will repeat the process for 'sampy', the other normal dist vector, and check the outcome

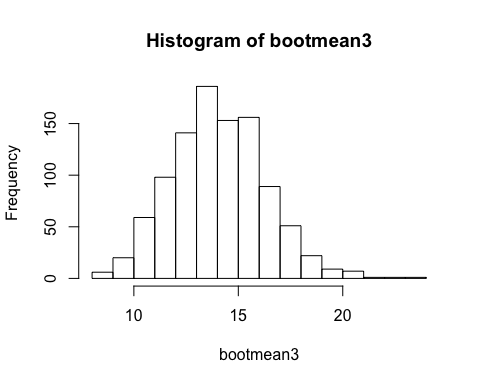
bootmean2 <- numeric(R)  
for (i in 1:R) {  
 bootsample2 <- sample(sampy, size=length(sampy), replace=TRUE)  
 bootmean2[i] <- mean(bootsample2)}  
  
hist(bootmean2)



# Again, a normal distribution is realized.

# Next, I will repeat the procedures for both of the exponential dist vectors 'xampx' and 'xampy'

bootmean3 <- numeric(R)  
for (i in 1:R){  
 bootsample3 <- sample(xampx, size=length(xampx), replace=TRUE)  
 bootmean3[i] <- mean(bootsample3)}  
  
hist(bootmean3)



# Here are the results from the vector 'xampx', which is an exponential distribution.

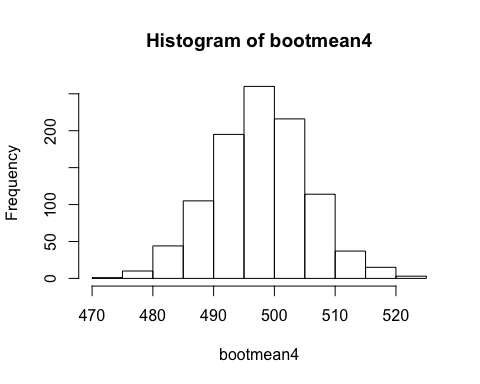
# We can see that the histogram approaches normality

# This helps visualize the concept of central limit theorem

# Even though our original data set was of exponential dist, when we sampled from that over and over, the plotted means produce a histogram that approaches normality.

# Now, I will show this concept again with the other exponential dist vector 'xampy'

bootmean4 <- numeric(R)  
for (i in 1:R){  
 bootsample4 <- sample(xampy, size=length(xampy), replace=TRUE)  
 bootmean4[i] <- mean(bootsample4)}  
  
hist(bootmean4)



# bootstrap success!

# again, we see that the results produce a histogram that approaches normality. These two histograms are quite different, and shows how regardless of the distribution of the original data, bootstrapping will yield a normal dist

attach(mtcars)  
par(mfrow=c(1,2))  
hist(xampy)  
hist(bootmean4)

