Question 4

Joseph Froelicher

9/24/2020

# Parts A & B

# individual sample size  
n <- 10  
  
# the number of simulations per sample size  
N <- 500  
  
# probability of success  
P <- 0.15  
  
# number of columns in the matrices  
C <- 5  
  
sample\_mean <- matrix(NA, N, C)  
  
iter <- 0  
  
for (n in seq(10, 50, 10)) { # 1:5 == seq(1,5,1)  
 iter <- iter + 1  
  
 for (i in 1:N) {  
 draws <- rbinom(n, 1, P)  
 sample\_mean[i, iter] <- mean(draws)  
 }  
}  
  
head(sample\_mean)

## [,1] [,2] [,3] [,4] [,5]  
## [1,] 0.1 0.15 0.13333333 0.125 0.22  
## [2,] 0.1 0.00 0.10000000 0.100 0.12  
## [3,] 0.2 0.25 0.06666667 0.150 0.16  
## [4,] 0.2 0.30 0.16666667 0.075 0.06  
## [5,] 0.0 0.05 0.20000000 0.250 0.14  
## [6,] 0.2 0.15 0.10000000 0.200 0.20

# Part C

means\_table <- data.frame(  
 mean(sample\_mean[,1]),  
 mean(sample\_mean[,2]),  
 mean(sample\_mean[,3]),  
 mean(sample\_mean[,4]),  
 mean(sample\_mean[,5])  
)  
  
colnames(means\_table) <- c('n = 10', 'n = 20', 'n = 30', 'n = 40', 'n = 50')  
means\_table

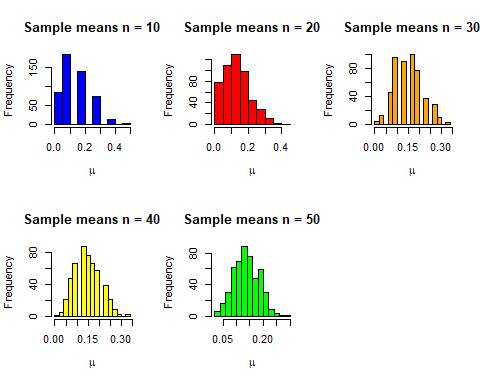
## n = 10 n = 20 n = 30 n = 40 n = 50  
## 1 0.1522 0.154 0.1537333 0.1492 0.14668

sd\_table <- data.frame(  
 sd(sample\_mean[,1]),  
 sd(sample\_mean[,2]),  
 sd(sample\_mean[,3]),  
 sd(sample\_mean[,4]),  
 sd(sample\_mean[,5])  
)  
  
colnames(sd\_table) <- c('n = 10', 'n = 20', 'n = 30', 'n = 40', 'n = 50')  
sd\_table

## n = 10 n = 20 n = 30 n = 40 n = 50  
## 1 0.1069488 0.08134639 0.0637772 0.05810432 0.04709792

# Part D

par(mfrow = c(2,3))  
  
hist(  
 sample\_mean[,1],  
 breaks = 15,  
 main = 'Sample means n = 10',  
 xlab = expression(mu),  
 col = 'blue'  
)  
  
hist(  
 sample\_mean[,2],  
 breaks = 15,  
 main = 'Sample means n = 20',  
 xlab = expression(mu),  
 col = 'red'  
)  
  
hist(  
 sample\_mean[,3],  
 breaks = 15,  
 main = 'Sample means n = 30',  
 xlab = expression(mu),  
 col = 'orange'  
)  
  
hist(  
 sample\_mean[,4],  
 breaks = 15,  
 main = 'Sample means n = 40',  
 xlab = expression(mu),  
 col = 'yellow'  
)  
  
hist(  
 sample\_mean[,5],  
 breaks = 15,  
 main = 'Sample means n = 50',  
 xlab = expression(mu),  
 col = 'green'  
)



# Part E

There is probably an argument to be made that the values begin to look normal at the earliest at . However, I would say that not until is where we can really see visually the resemblence of a normal curve, and at the sample means surely resemble a normal curve.