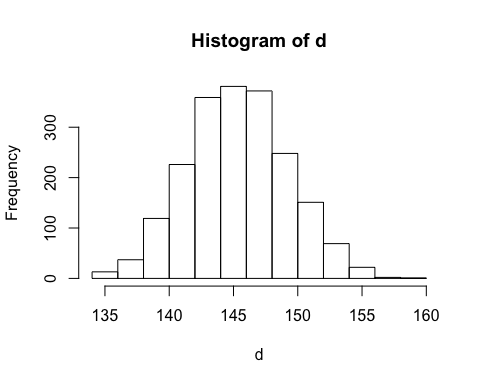
note05

Matts966

2018/01/30

# 1

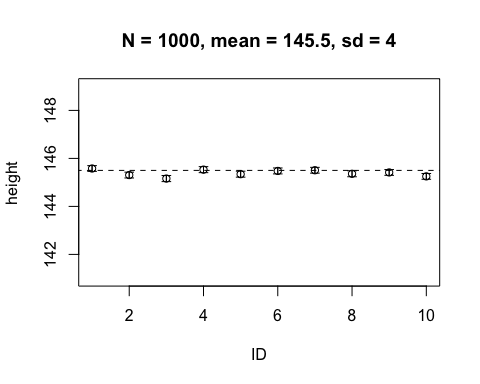
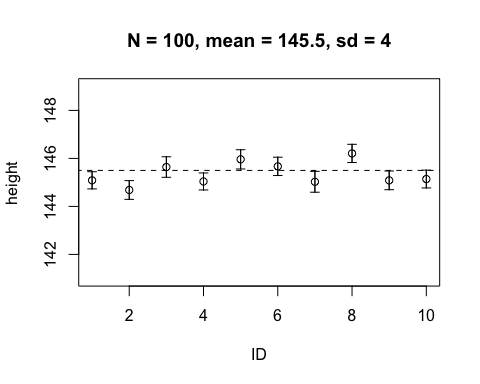
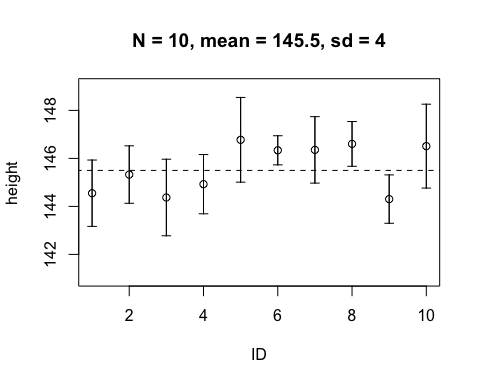
# 母集団を生成  
set.seed(131)  
d <- rnorm(2000, mean=145.5, sd=4.0)  
hist(d)



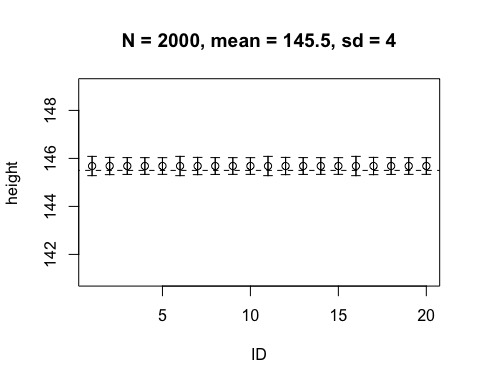
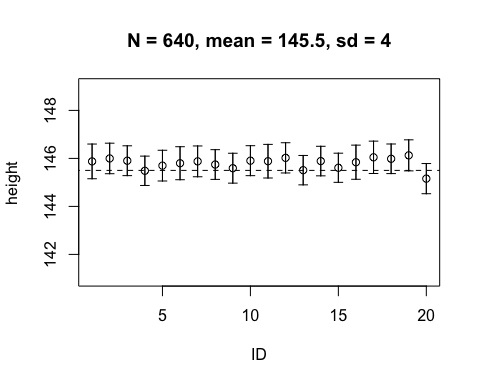
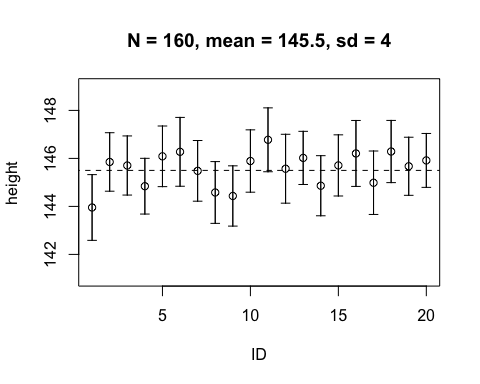
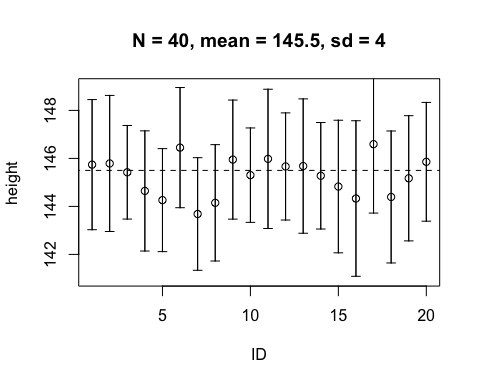
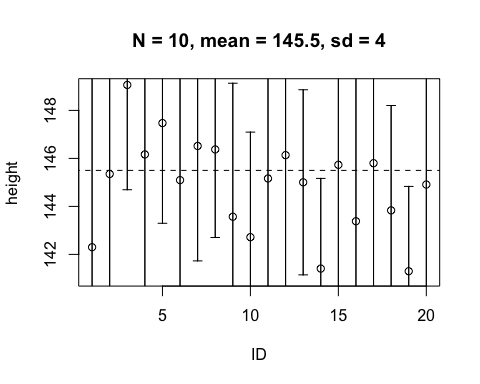
mean(d)

## [1] 145.3573

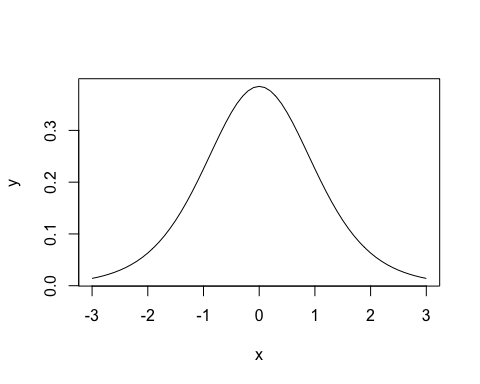
# for文を使ってサンプル数を増やしてみる。  
# 標準誤差が小さくなり、平均値も母集団のものに近づいていった。  
# はじめsample関数は繰り返し使うと次に使うときも母集団が減った状態でスタートするのがデフォルトなのかと思ったが、replaceのT, Fをかえて実験したところ、関数1回分で母集団を超えない限りエラーが出なかったので、関数内でのみ抽選済みか否かを保持しているのだ、と気づいた。僕に取っては少し勘違いしやすいところでした。  
N = c(10, 100, 1000)  
for (n in N) {  
 mean\_x <- c()  
 sd\_x <- c()  
 se\_x <- c()  
 for (i in 1:10) {  
 x <- sample(d, n)  
 mean\_x[i] <- mean(x)  
 sd\_x[i] <- sqrt(var(x))  
 se\_x[i] <- sd\_x[i] / sqrt(n)  
 }  
 xaxis <- seq(1, 10, 1)  
 plot(xaxis, mean\_x, xlim = c(1, 10), ylim = c(141, 149), xlab = "ID", ylab = "height", main=paste("N = ", n, ", mean = 145.5, sd = 4", sep=""))  
 arrows(xaxis, mean\_x + se\_x, xaxis, mean\_x - se\_x, angle = 90, length = 0.05)  
 arrows(xaxis, mean\_x - se\_x, xaxis, mean\_x + se\_x, angle = 90, length = 0.05)  
 lines(c(0, 11), c(145.5, 145.5), lty=2)  
}

 【考察】 標準誤差に関してもサンプル数を多くしてみた。for文を使うことで、同時に図を生成することができた。sample関数に関しては、replaceは関数のコーリングの際のみ値が保持されるようなので、Falseのままでうまくいっている。

d <- rnorm(2000, mean = 145.5, sd = 8)  
  
# for文を使ってサンプル数を増やしてみる。  
# 標準誤差が小さくなることでだんだんと95％信頼区間が毋平均が外れなくなっていく  
N <- c(10, 40, 160, 640, 2000)  
for (n in N) {  
 mean\_x <- c()  
 sd\_x <- c()  
 se\_x <- c()  
 for (i in 1:20) {  
 x <- sample(d, n)  
 mean\_x[i] <- mean(x)  
 sd\_x[i] <- sqrt(var(x))  
 se\_x[i] <- sd\_x[i] / sqrt(n)  
 }  
 xaxis <- seq(1, 20, 1)  
 plot(xaxis, mean\_x, xlim = c(1, 20), ylim = c(141, 149), xlab = "ID", ylab = "height", main=paste("N = ", n, ", mean = 145.5, sd = 4", sep=""))  
 arrows(xaxis, mean\_x - se\_x \* qt(1 - 0.025, df = N - 1), xaxis, mean\_x + se\_x \* qt(1-0.025, df = N - 1), angle = 90, length = 0.05)  
 arrows(xaxis, mean\_x + se\_x \* qt(1 - 0.025, df = N - 1), xaxis, mean\_x - se\_x \* qt(1-0.025, df = N - 1), angle = 90, length = 0.05)  
 lines(c(0, 21), c(145.5, 145.5), lty=2)  
}

 【考察】 N = 2000でもエラーが出ておらず、sample関数に関しては、replaceはFalseのままでうまくいっていることを示している。標準誤差が小さくなることでだんだんと95％信頼区間が毋平均が外れなくなっていく様子をしっかりと実感できる内容だった。 \* 4章 演習問題第3問

# 自由度7のt分布  
x <- seq(-3, 3, 0.1)  
y <- dt(x, df = 7)  
plot(x, y, type = 'l')



# 上側のt値、下側のt値  
qt(1 - 0.025, df = 7)

## [1] 2.364624

qt(0.025, df = 7)

## [1] -2.364624

【考察】 curve関数をもちいる方法もあるようだったが、素直にｘ軸ｙ軸を計算して配列に格納し、プロットした。t値に関しても教科書通り求めることができた。