

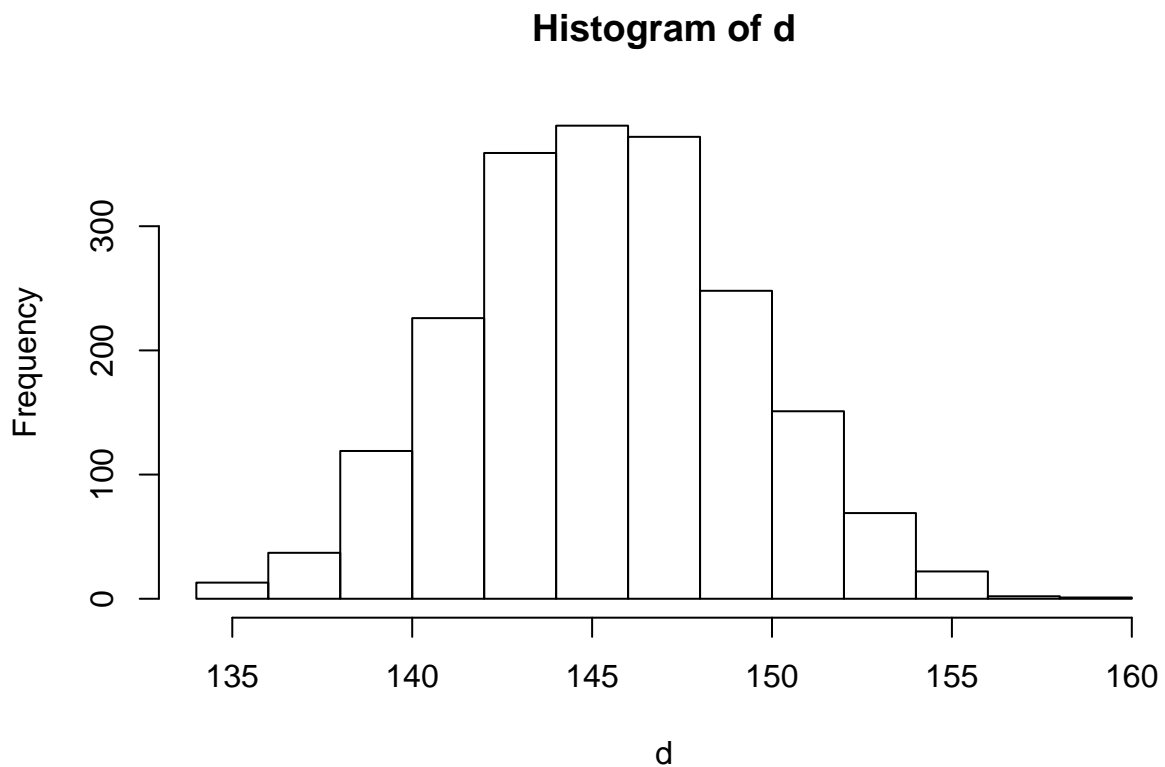
# note05

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
```

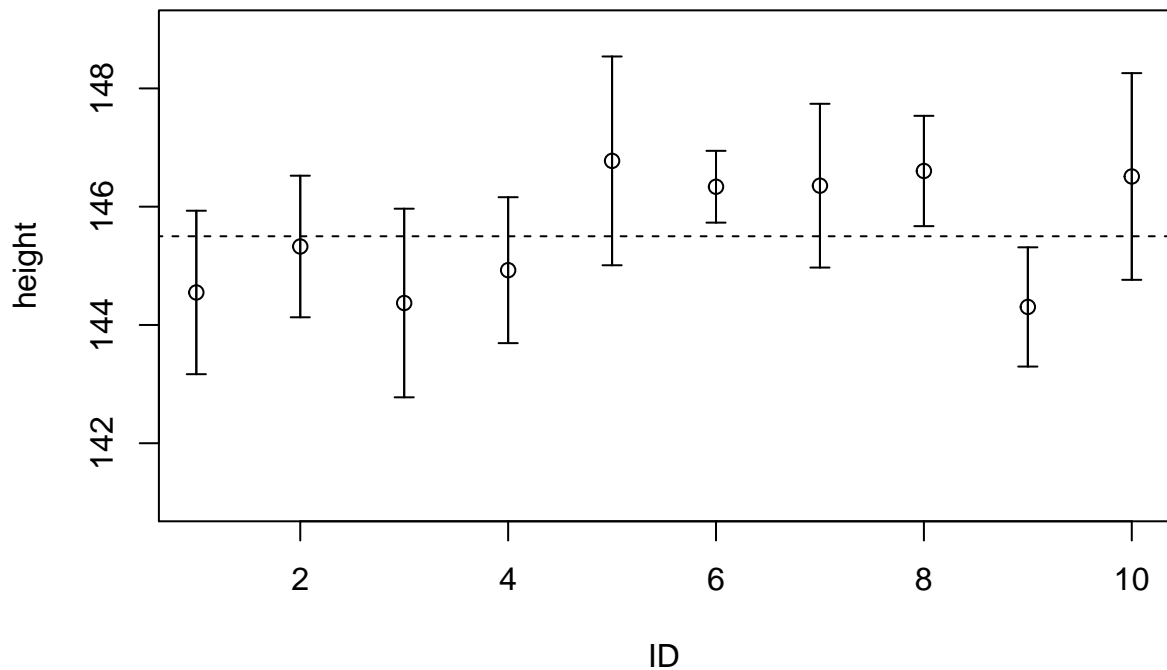
```
# for文を使ってサンプル数を増やしてみる。
# 標準偏差が小さくなり、平均値も母集団のものに近づいていった。
# はじめ sample 関数は繰り返し使うと次に使うときも母集団が減った状態でスタートするのがデフォルトなのかと思ったが、
N = c(10, 100, 1000)
for (n in N) {
  mean_x <- c()
  sd_x <- c()
}
```

```

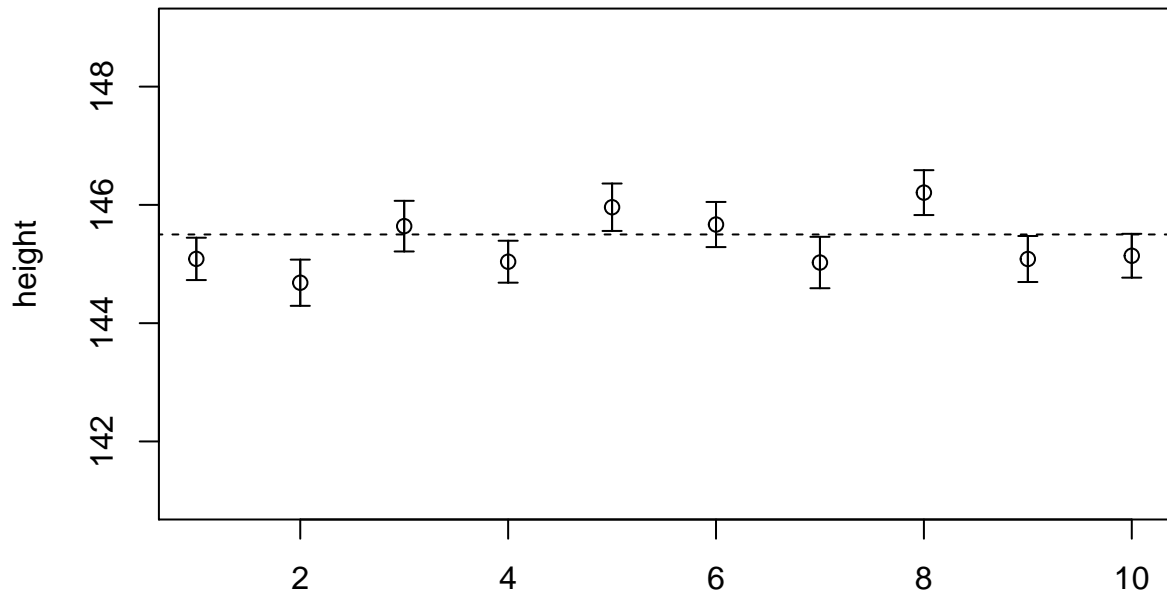
se_x <- c()
for (i in 1:10) {
  x <- sample(d, n)
  mean_x[i] <- mean(x)
  sd_x[i] <- sqrt(var(x))
  se_x[i] <- sd_x[i] / sqrt(n)
}
xaxis <- seq(1, 10, 1)
plot(xaxis, mean_x, xlim = c(1, 10), ylim = c(141, 149), xlab = "ID", ylab = "height", main=paste(
arrows(xaxis, mean_x + se_x, xaxis, mean_x - se_x, 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)
}

```

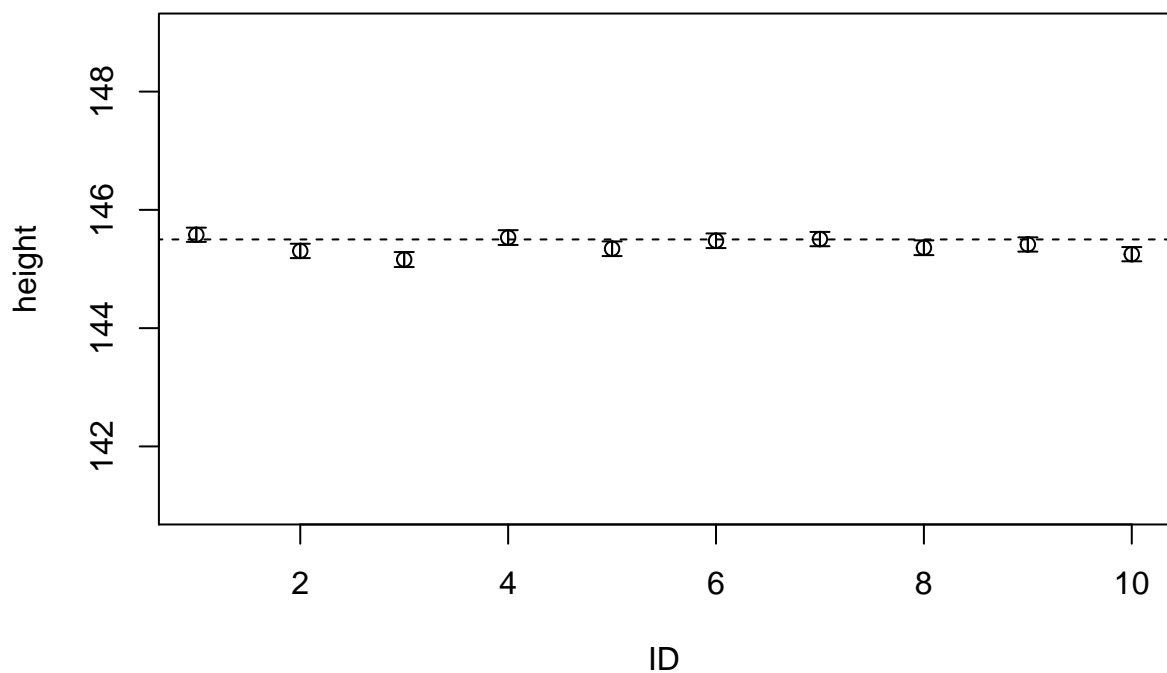
**N = 10, mean = 145.5, sd = 4**



**N = 100, mean = 145.5, sd = 4**



**N = 1000, mean = 145.5, sd = 4**



【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。
```

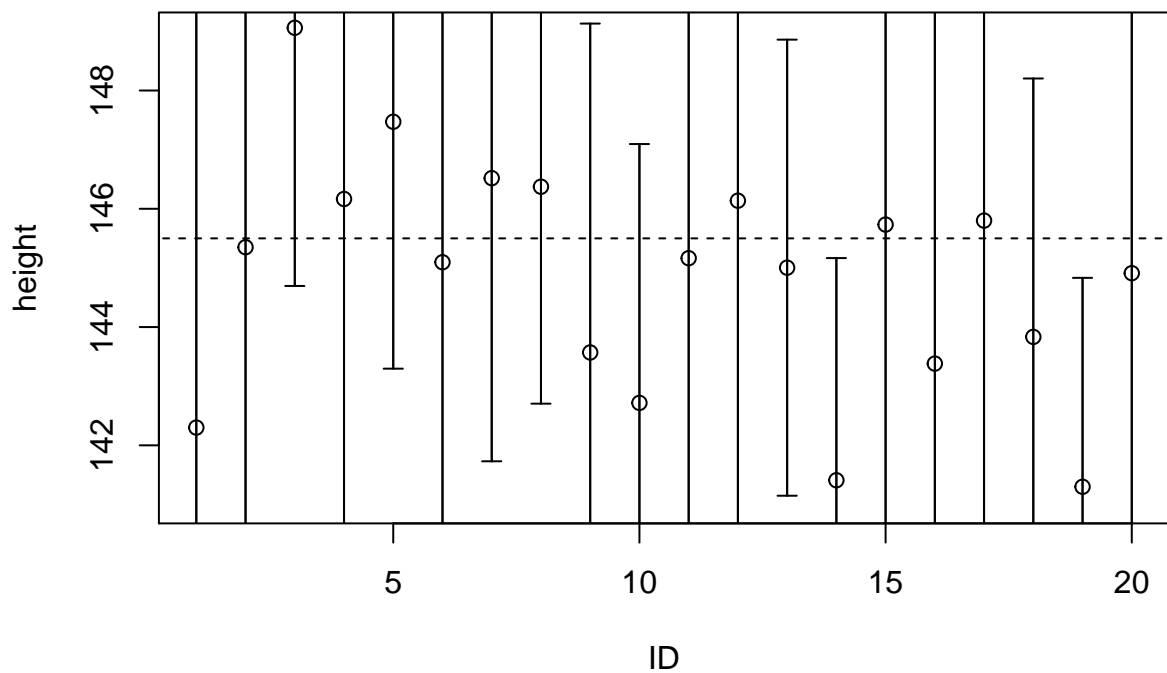
```
# 標準誤差が小さくなることでだんだんと 95 %信頼区間が母平均が外れなくなっていく
```

```

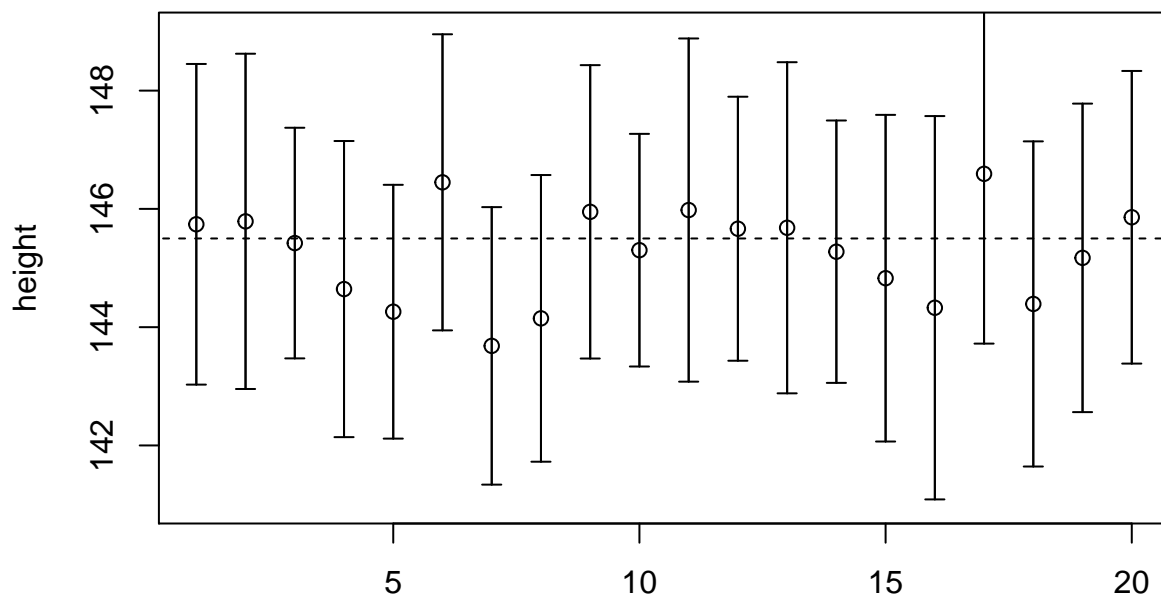
N <- c(10, 40, 160, 640, 2000)
for (n in N) {
  mean_x <- c()
  sd_x <- c()
  se_x <- c()
  for (i in 1:20) {
    x <- sample(d, n)
    mean_x[i] <- mean(x)
    sd_x[i] <- sqrt(var(x))
    se_x[i] <- sd_x[i] / sqrt(n)
  }
  xaxis <- seq(1, 20, 1)
  plot(xaxis, mean_x, xlim = c(1, 20), ylim = c(141, 149), xlab = "ID", ylab = "height", main=paste(
arrows(xaxis, mean_x - se_x * qt(1 - 0.025, df = N - 1), xaxis, mean_x + se_x * qt(1-0.025, df = N
arrows(xaxis, mean_x + se_x * qt(1 - 0.025, df = N - 1), xaxis, mean_x - se_x * qt(1-0.025, df = N
lines(c(0, 21), c(145.5, 145.5), lty=2)
}

```

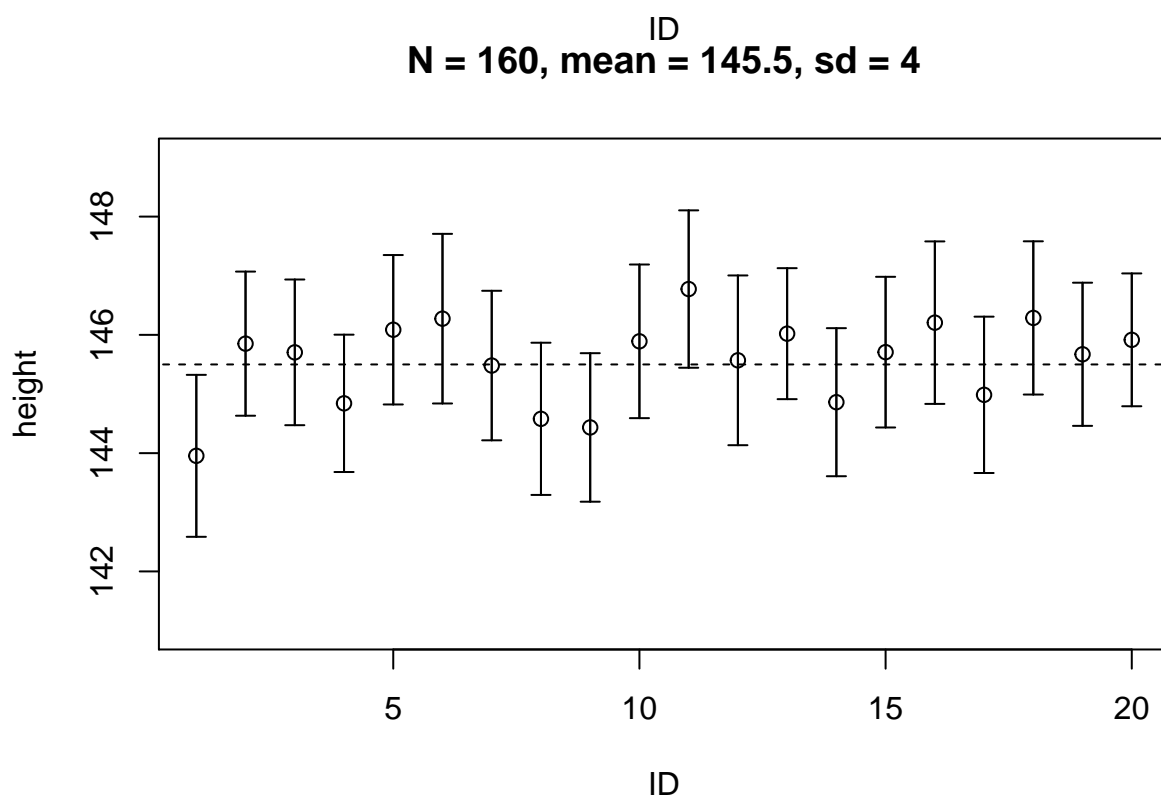
**N = 10, mean = 145.5, sd = 4**



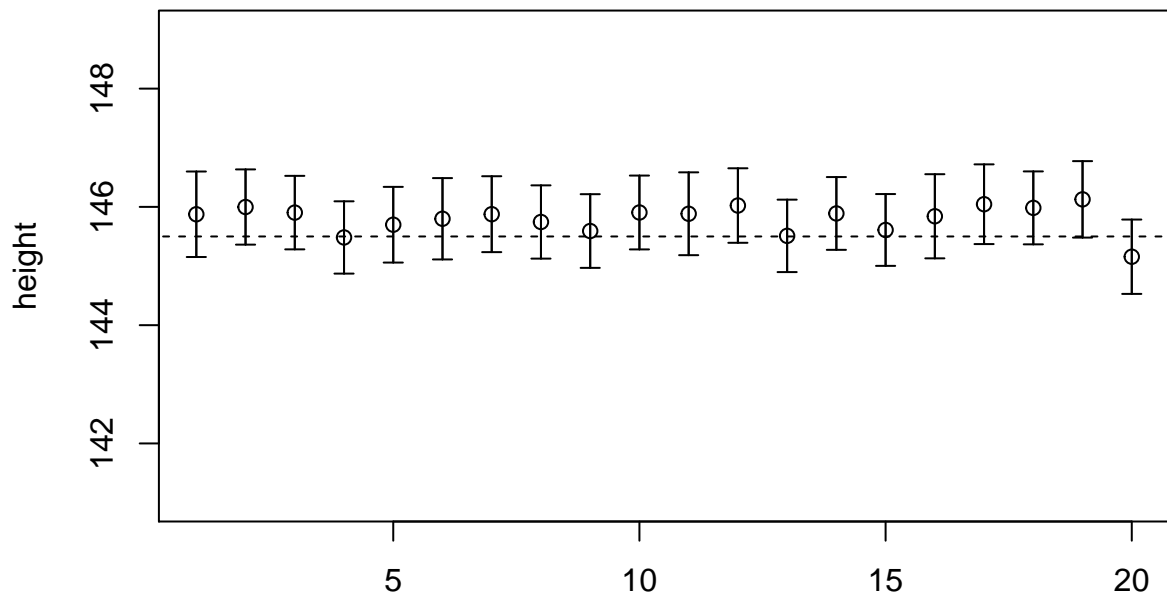
**N = 40, mean = 145.5, sd = 4**



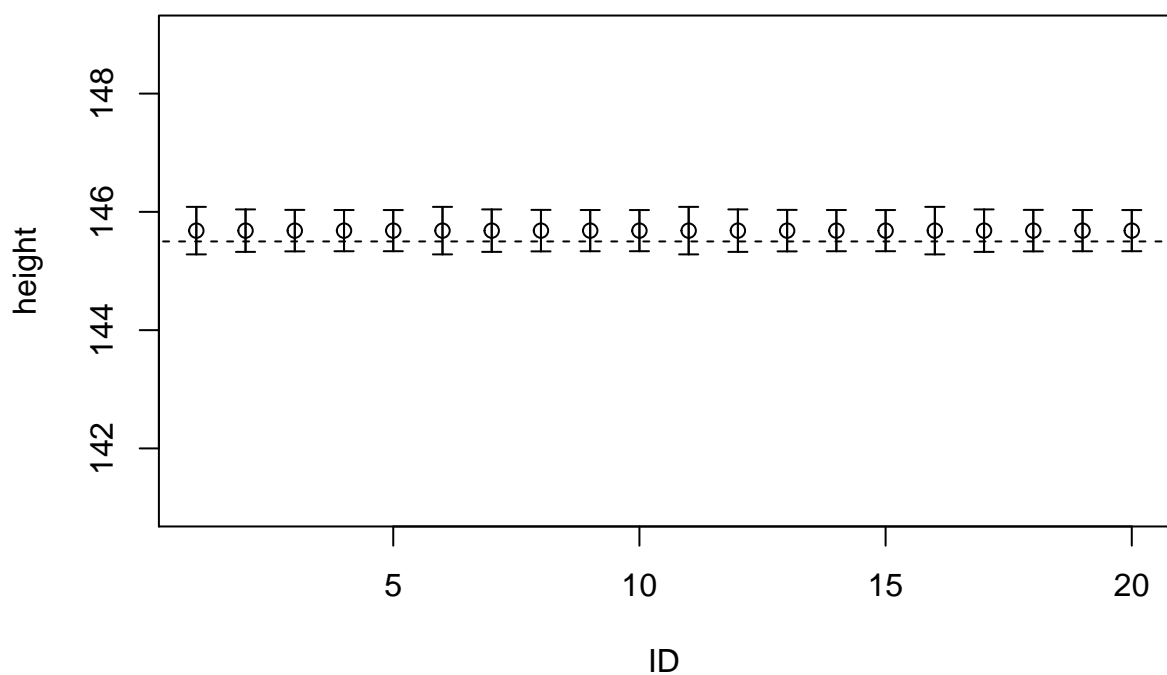
**N = 160, mean = 145.5, sd = 4**



**N = 640, mean = 145.5, sd = 4**

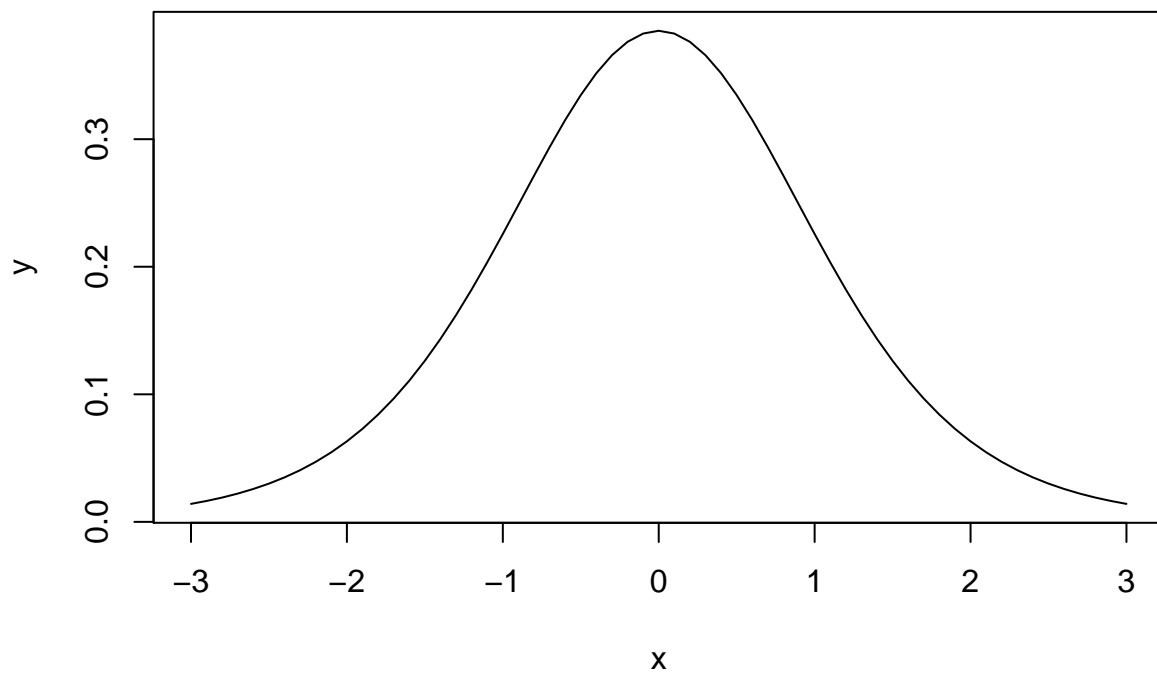


**N = 2000, mean = 145.5, sd = 4**



【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 %信頼区間が母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 関数を持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。