

Biostatistic homework 1

Qiang Liu

2022-10-6

First include packages and read the data in the .txt file

```
> library(moments)
> exercise_data <- scan('homework_exercise1_data.txt')
```

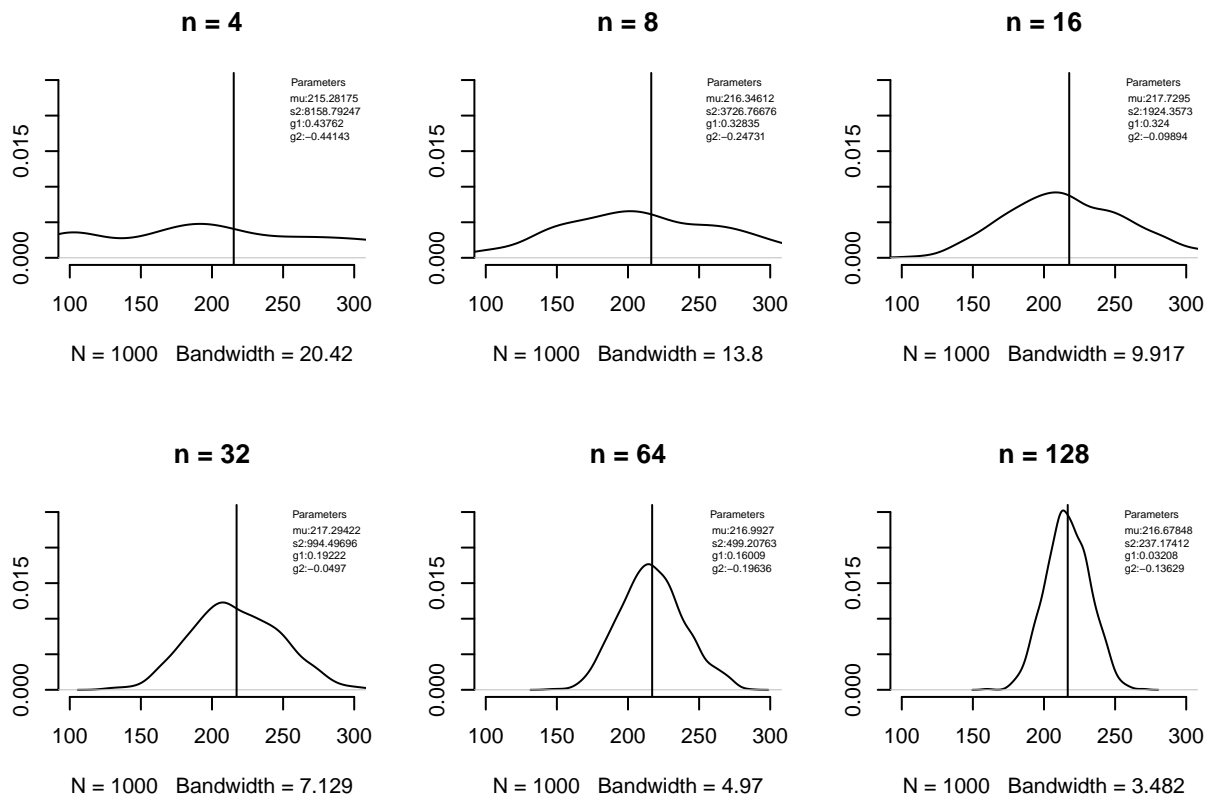
Plot of different sample size ($n = 4, 8, 16, 32, 64, 128$)

```
> exercise_mean <- mean(exercise_data)
> N <- 1000
> # 修改图参数，一排三个图，放两排
> opar <- par(no.readonly = TRUE)
> par(mfrow = c(2,3))
> par(pin = c(1.6,1))
> # 选取 n 个样本，计算对应的参数并放置于右上，将样本数据按 density 分布作图，并以均值作为 abline
> exercise_function <- function(n){
+   x <- unlist(lapply(1:N,function(i){
+     mean(sample(exercise_data,size = n,replace = FALSE))
+   }))
+   out <- c(mu=mean(x),s2 = var(x),g1 = skewness(x),g2 = kurtosis(x)-3)
+   plot(density(x),bty = 'n',xlim = c(100,300),ylim = c(0,0.025),
+     main=sprintf("n = %s",n))
+   abline(v = mean(x))
```

```

+   lgd <- sprintf("%s:%s",names(out),round(out,5))
+   legend('topright',legend = lgd,ncol = 1,box.lwd = NA,cex = 0.5,title = "Parameters")
+   out
+ }
> # 选取 n 为 4, 8, 16, 32, 64, 128 的每种情况画图
> for (i in c(4,8,16,32,64,128)){
+   exercise_function(i)
+ }

```



```

> par<-opar

```

从图中可以看出，整体而言当 n 选取变大的时候， μ 值几乎不变， s^2 值越变越小， g_1 和 g_2 值也整体变小，但由于样本选取量的随机性，可能 g_1 值和 g_2 值会有所不同，整体而言，当 n 变大时，峰度值更接近于 0，此时的总体数据分布与正态分布的陡缓程度相似，同时偏度值也接近于 0，此时的总体数据分布与正态分布的偏斜程度也相似，因此对于该样本而言，选取样本量增加，总体数据分布更接近于正态分布。

Histogram of exercise_data

```
> # 作样本数据的 histogram 图并作 line, 将 histogram 颜色改为白色并删去 x 轴名字和图表名
> exercise_den <- density(exercise_data)
> par(pin = c(4,2))
> hist(exercise_data,freq = FALSE,breaks = 40,xlab = ' ',
+       ylab = 'Density',main = ' ',col = 'white')
> lines(exercise_den,col = 'red',lwd = 1.5)
```

