Advanced Statistical Method Hw 2

Dohyup Shin

2021-10-13

Exercise 3.4

(a) Run the following simulation 200 times:
- $x_i \overset{i.i.d}{\sim} N(\mu_i, 1)$ for $i = 1, 2, \ldots, 500$
- $\mu_i = 3i/500$
- i_{max} = index of largest x_i
- $d = x_{i_{max}} - \mu_{i_{max}}$

(b) Plot the histogram of the 200 d values

(c) What is the relation to Figure 3.4?

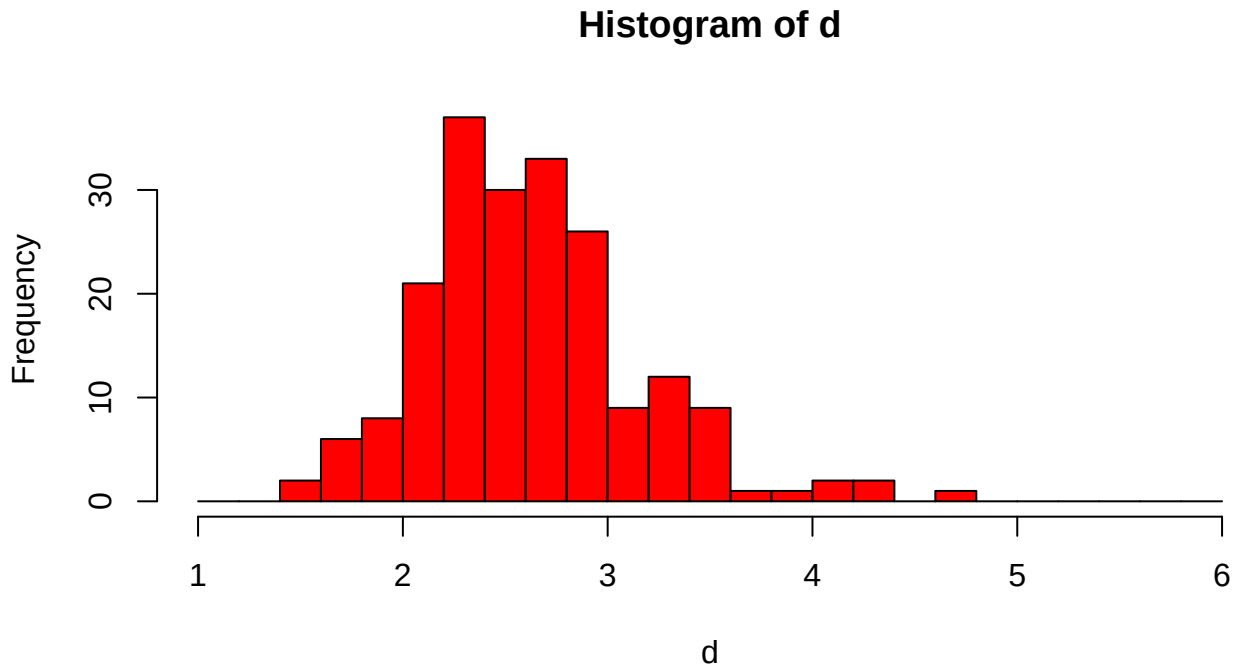
Solution

(a)

```
#mean vector
mu <- seq(0.006, 3, 0.006)

#x[i,j] ~ N(mu[j], 1), 1=< i =< 200, 1=< j =< 500
x <- matrix(NA, nrow = 200, ncol = 500)

#the index of the largest x_i in the i_th simulation
idx <- rep(NA, 200)

#the difference between the largest x_i and the mean at that time in the i-th simulation
d <- rep(NA, 200)


#simulating 200 times
for(i in 1:200){
  for(j in 1:500){
    x[i,j] <- rnorm(1, mu[j], 1) #Generating one random number in the N(mu[j], 1)
  }
  idx[i] <- which.max(x[i,]) #Find the largest element value in the x[i,]
  d[i] <- x[i, idx[i]] - mu[idx[i]]
}
```

(b)

```
hist(d, breaks = seq(1, 6, 0.2), col = 'red')
```



(c) In Figure 3.4, the histogram of the x_i values does in fact reveal some large values, $x_{610} = 5.29$ which is distributed $N(\mu_{610}, 1)$. Then, can we say that x_{610} is a good estimate for μ_{610}?. Even though x_{610} was unbiased for μ_{610}, we would worry that focusing attention on the largest of 6033 values would produce an upward bias. So we can expect that our estimate should downwardly correct 5.29. this phenomenon is reffered to as selection bias. Theoretically, $x_{i_{max}} - \mu_{i_{max}}$ is distributed $N(0, 1)$ for each simulation, but by above simulation, we show that $x_{i_{max}} - \mu_{i_{max}}$ was always positive for every simulation. This means that $x_{i_{max}}$ is an upwardly biased estimate for $\mu_{i_{max}}$. Therefore, by above result, estimated value $x_{i_{max}}$ for $\mu_{i_{max}}$ should be modified downward.