Bootstrap Lab

Shaoran Sun

May 3, 2016

Estimating a population mean

1

The true mean of the population is **10**.

The estimate from the sample is **10.08782**.

It is NOT exactly correct, but very close.

2

The standard deviation of the sample is **2.865351**.

The estimate from #1 is within 2 of 2.865351 from the true mean.

3

From the histogram and the calculation, **95.1%** of the time the sample mean is within 2 standard deviation of the true mean.

4

The 95% confidence interval for the true population mean is **(8.6329, 10.92518)**, the true mean 10 is indeed within this interval.

The bootstrap

5

The values in the sample from #1 are:

12.283483 5.706046 11.856188 10.634344 15.336947 6.619884 12.313170 8.376584 12.408567 15.615882 12.277312 6.892050 10.494954 8.182230 7.245307 10.228124 8.320412 8.981784 4.659279 10.500860 6.711899 8.772800 7.221565 12.602552 10.233796

The mean of the sample is **9.779041**.

6

There are duplicates in the bootstrap sample.

The mean is **9.424798**.

7

From the second histogram, the second one is more similar to normal, compared to the one in #3.

8

The mean of bootmeans is **9.757891**, and the standard deviation is 0.5751015. Standard Error is **0.1150203** The 95% confidence interval is **(9.52785, 9.987932)**

A real example - backpack weights

9

The sample size is 20.

From the third histogram, it does not appear to be normally distributed. So we are NOT able to use the precedure in #4 to find a confidence interval.

10

$SE(\bar{X}) = \frac{S}{\sqrt{n}} = \frac{8.47209}{\sqrt{20}} = 1.894417$

$SE(\bar{X})$ is 1.894417. Using bootstrap, the standard error is 1.867714. These two are very close.

11

The mean of the sample is 13.25. With 95% confidence interval, we have the interval of (9.514572, 16.98543).

Using method in #4, we have the interval, with 95% confidence, (9.461166, 17.03883).

Method in #4 gives us a wider interval.

Lab Summary

The 95% bootstrap interval is (9.514572, 16.98543). 95% of the chances that the true mean falls in this interval.

The 95% confidence interval is (9.461166, 17.03883).

The two confidence interval is pretty close, the bootstrap interval is more narrow than the traditional confidence interval.

Appendix

1

```
mysample = rnorm(25, 10, 3)
mean(mysample)
```

```
## [1] 10.67073
```

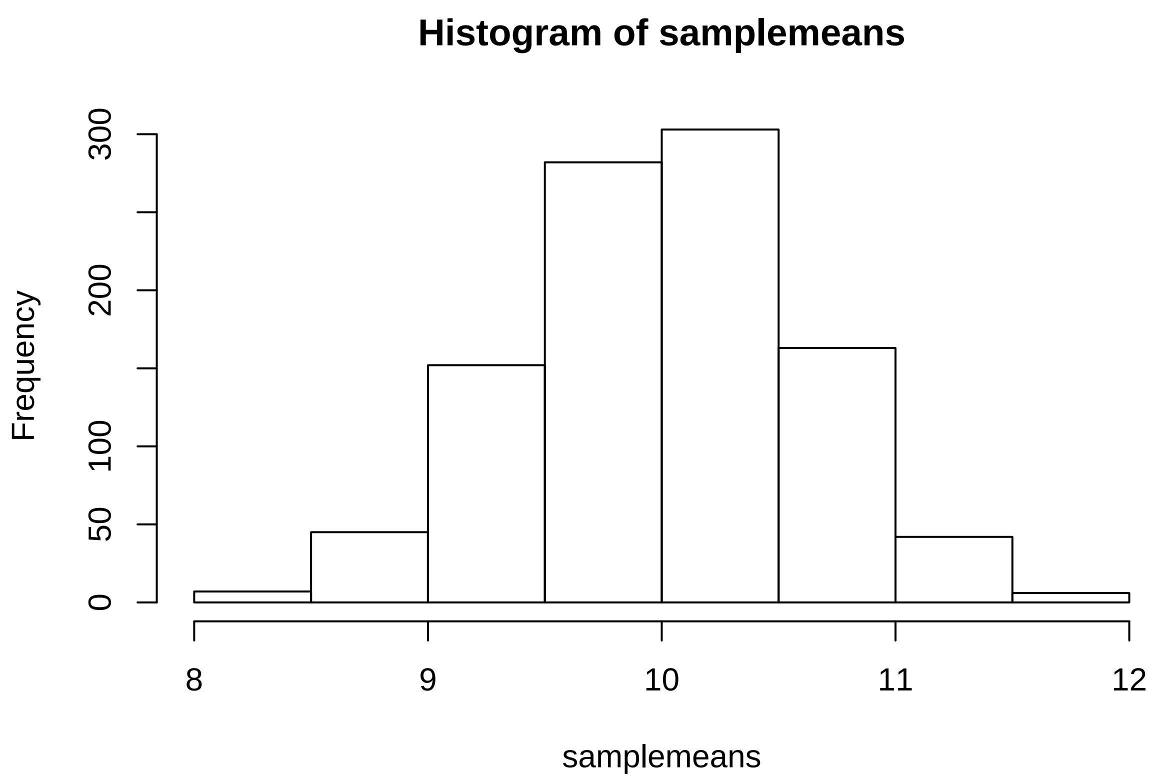
2

```
mysample = rnorm(25, 10, 3)
sd(mysample)
```

```
## [1] 3.150479
```

Histogram

```
samplemeans = rep(NA,1000)
for (i in 1:1000) {
mysample.i = rnorm(25, 10, 3)
samplemeans[i] = mean(mysample.i)
}
hist(samplemeans)
```



```
sum(samplemeans >= 10-2*3/sqrt(25) & samplemeans <= 10+2*3/sqrt(25))/1000
```

```
## [1] 0.951
```

4

```
x.bar = mean(mysample)
s = sd(mysample)
x.bar - 2*s/sqrt(25)
```

```
## [1] 8.223215
```

```
x.bar + 2*s/sqrt(25)
```

```
## [1] 10.7436
```

5

```
mysample
```

```
##  [1]  7.400178  4.044496 12.758516 13.448860 12.037603 11.211569  4.628307
##  [8]  9.566486 11.611713  9.637955  9.336091  8.064564  8.110408  8.412179
## [15] 12.654044 13.745643 11.004314  2.934610  9.516066  9.416519 12.190208
## [22]  5.883679  4.719512 13.373955 11.377686
```

```
mean(mysample)
```

```
## [1] 9.483406
```

6

```
bootsample = sample(mysample, 25, replace=T)
bootsample
```

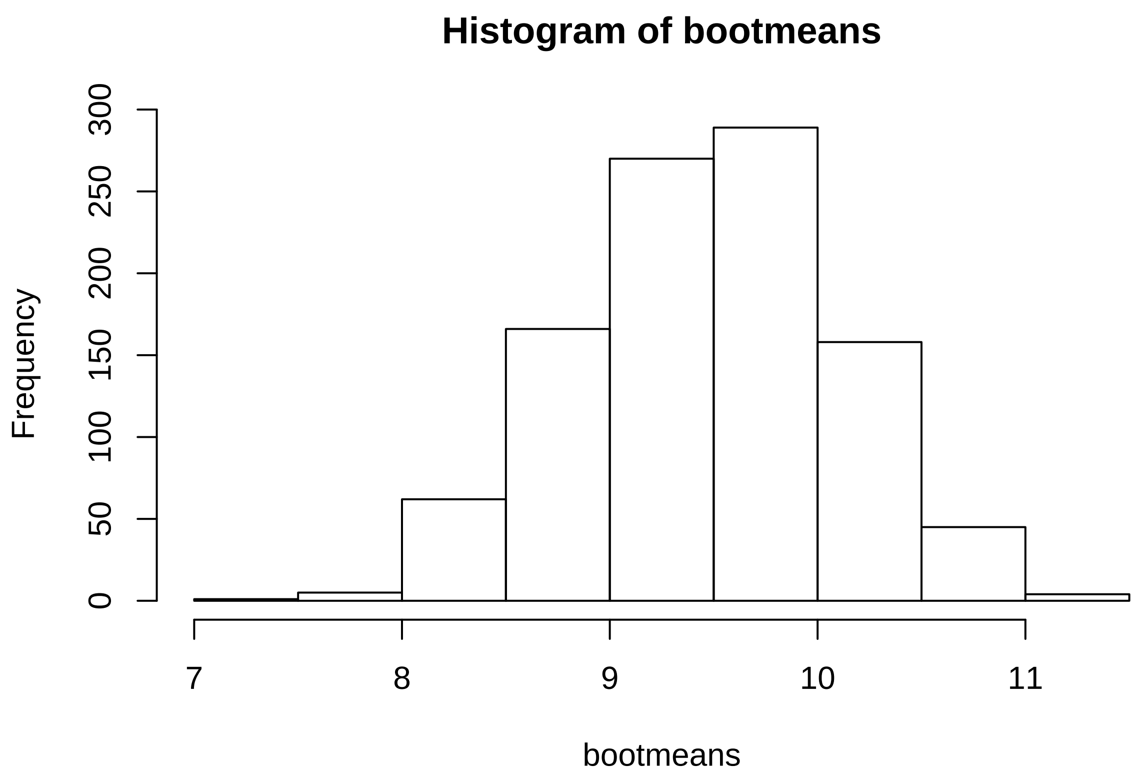
```
##  [1] 12.037603 11.377686 12.037603  9.516066  2.934610  2.934610 11.004314
##  [8] 13.373955  9.637955  8.064564 12.654044  2.934610  9.637955 12.190208
## [15]  9.516066 12.654044 13.448860 11.377686 13.373955  2.934610  2.934610
## [22] 13.745643 13.745643 13.745643 11.377686
```

```
bootmean = mean(bootsample)
bootmean
```

```
## [1] 9.967609
```

SecondHistogram

```
bootmeans = rep(NA, 1000)
for (i in 1:1000) {
bootsample = sample(mysample, 25, replace=T)
bootmeans[i] = mean(bootsample)
}
hist(bootmeans)
```



8

```
sd(bootmeans)
```

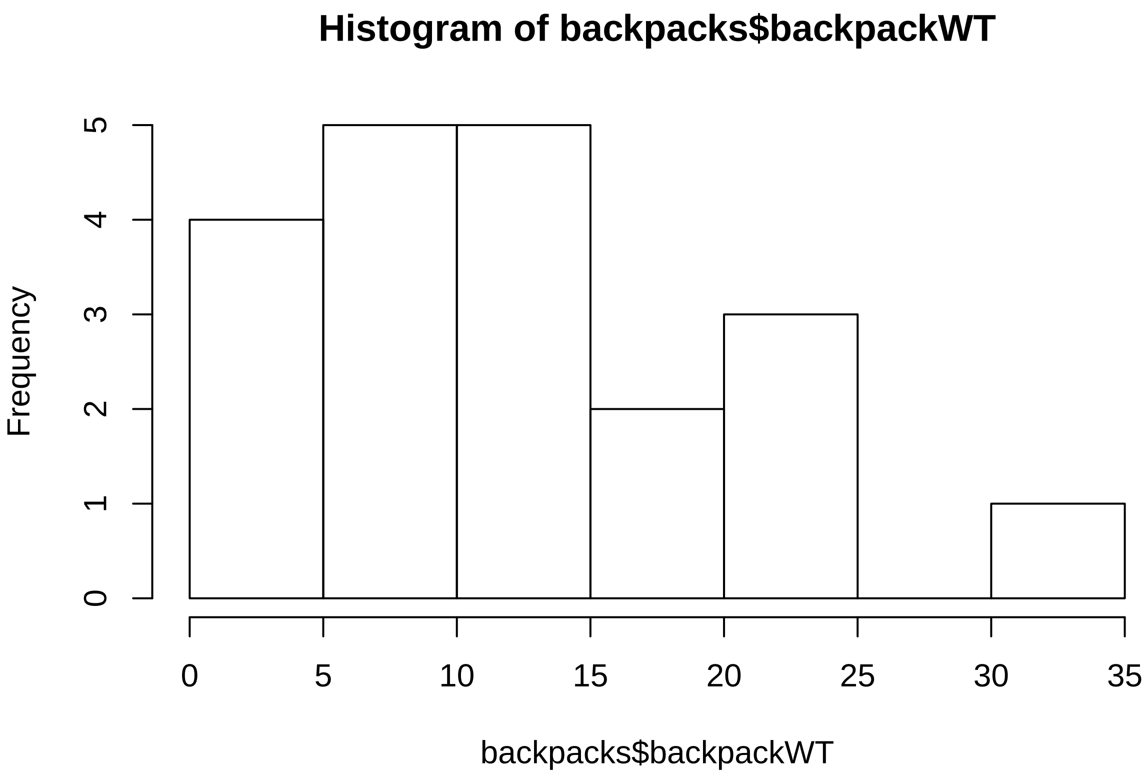
```
## [1] 0.6337749
```

```
mean(bootmeans)
```

```
## [1] 9.478446
```

ThirdHistogram

```
backpacks <- read.csv("~/Desktop/3480 Lab/backpacks.txt", sep="")
attach(backpacks)
hist(backpacks$backpackWT)
```




```r
sd(backpackWT)
```

```
## [1] 8.47209
```

10

```r
bootmeans = rep(NA, 1000)
for (i in 1:1000) {
bootsample = sample(backpackWT, 20, replace=T)
bootmeans[i] = mean(bootsample)
}
sd(bootmeans)
```

```
## [1] 1.816877
```

11

```r
mean(backpackWT)
```

```
## [1] 13.25
```

```r
x.bar = mean(backpackWT)
s = sd(backpackWT)
x.bar - 2*s/sqrt(20)
```

```
## [1] 9.461166
```

```r
x.bar + 2*s/sqrt(20)
```

```
## [1] 17.03883
```