Lab 9: Bootstrap

Frank Woodling

May 3, 2016

1.

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

```
## [1] 9.558755
```

The true mean is 10 as we set the random distribution to mean 10. The sample mean is 10.02622, which is very close. It would be closer to 10 with a greater sample size.

2.

```
3/sqrt(25)
```

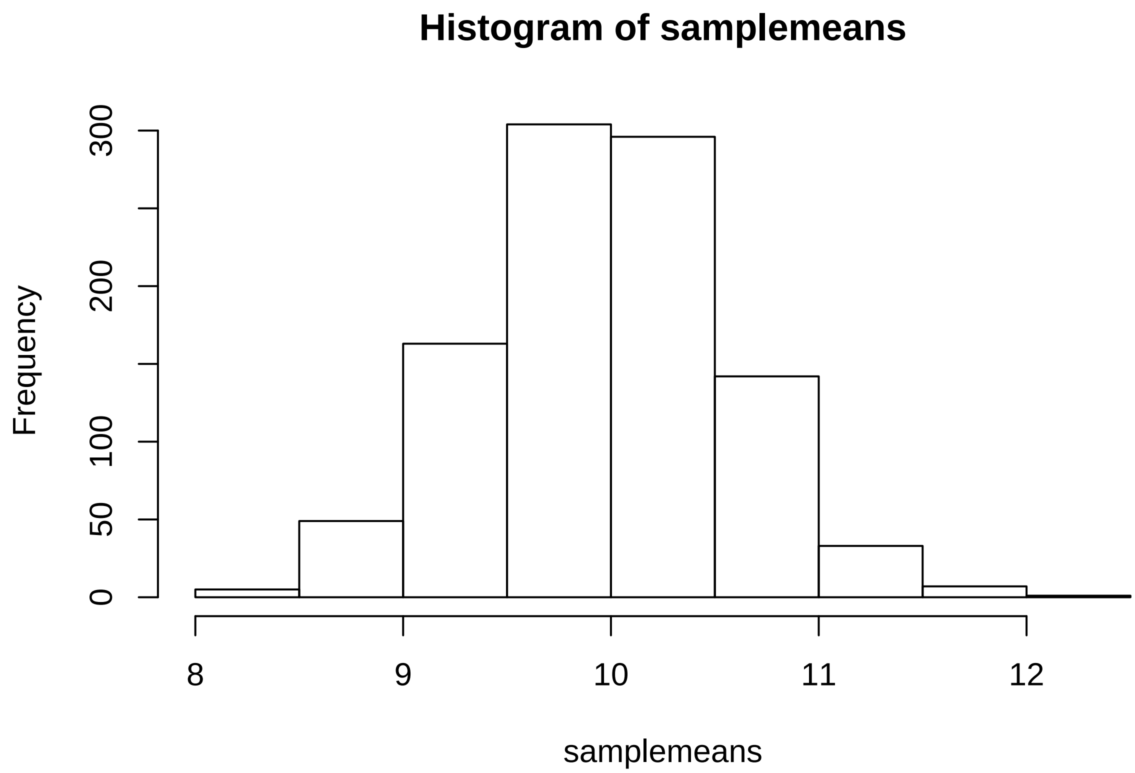
```
## [1] 0.6
```

The standard deviation of the sample mean is 0.6, so the estimate from part 1 is within a single standard deviation from the true mean.

3.

```
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 <= 0+2*3/sqrt(25))/1000
```

```
## [1] 0
```

It seems that all 1000 samples are within 2 standard deviations of the sample mean.

4.

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

```
## [1] 8.675752
```

```
## [1] 10.44176
```

The 95% CI is (8.82598, 11.22646).

5.

```r
mysample
```

```
##  [1]  6.923177 10.946754  9.933492  7.106787  7.922508  8.448396 12.319003
##  [8]  9.999012  7.745864  6.950265 12.922840  7.681069 11.785928  8.210820
## [15] 11.924036  9.961005 12.710195  7.269712 13.123680  7.611372  9.243249
## [22]  7.491645  8.448383 13.428443  8.861252
```

```r
mean(mysample)
```

```
## [1] 9.558755
```

The mean of the sample is 10.02622.

6.

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

```
##  [1]  9.243249  8.448383  8.448383  6.950265  8.210820 12.319003 13.428443
##  [8]  9.961005  9.961005 13.123680  9.243249  7.681069  9.961005  7.269712
## [15] 11.785928  9.961005  7.922508 11.785928  7.681069  7.681069  8.861252
## [22] 10.946754  7.922508  7.922508 13.428443
```

```r
bootmean = mean(bootsample); bootmean
```

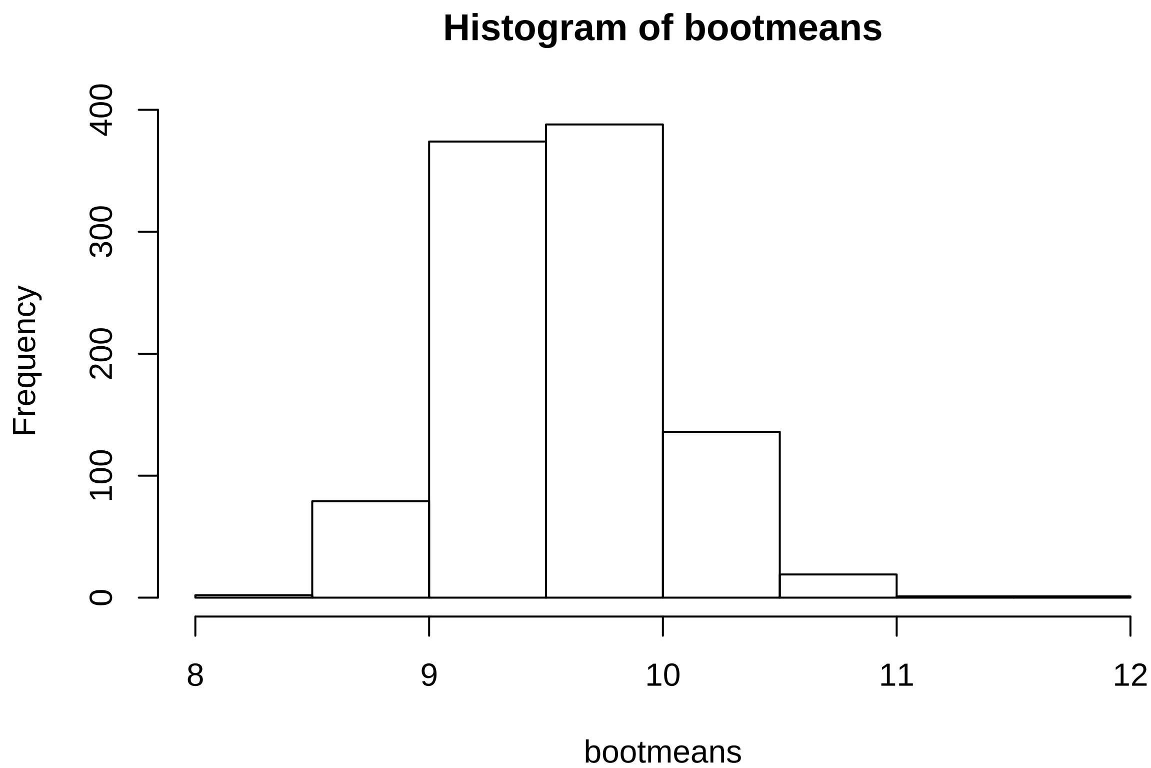
```
## [1] 9.60593
```

I see a couple duplicates. There are multiple instances of 10.141817 and 14.375842. That would explain why the bootstrap mean is higher than the original sample mean in part 5.

7.

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

hist(bootmeans)
```



The two histograms are very similar. You may be able to argue that the histogram in part 7 is slightly more normal or centered around a mean of 10.

8.

```
10-2*sd(bootmeans)
```

```
## [1] 9.129603
```

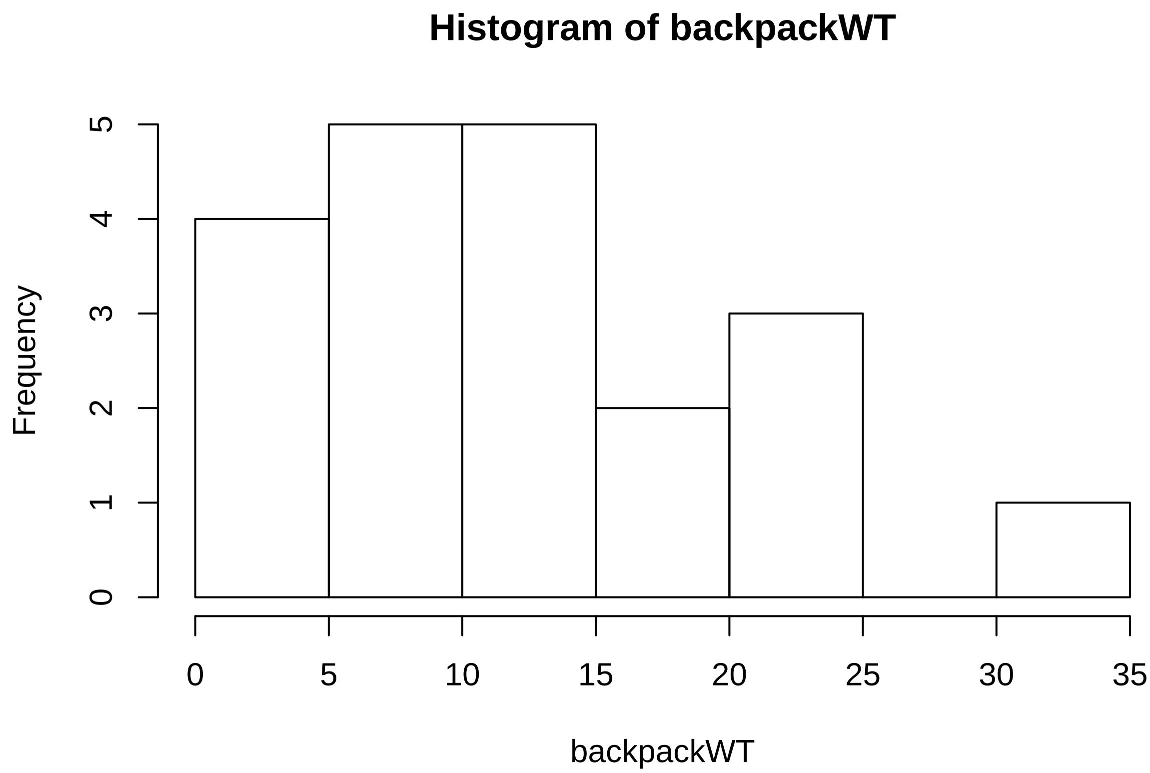
```
10+2*sd(bootmeans)
```

```
## [1] 10.8704
```

The 95% CI is (8.838112, 11.16189) compared to the interval in part 4 of (8.82598, 11.22646). The interval is slightly smaller in part 8, however, the two intervals are very close.

9.

```
data.2 <- read.csv("backpacks.txt", sep="")
attach(data.2)
hist(backpackWT)
```



The histogram is heavily right-skewed. We cannot use the method in number 4 since it is not normally distributed.

10.

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

```
## [1] 1.801378
```

```
sd(backpackWT)/sqrt(20)
```

```
## [1] 1.894417
```

The bootstrapped standard error is 1.848162, and the standard error using the formula is 1.894417.

11.

```r
mean(backpackWT)-2*sd(bootmeans)
```

```
## [1] 9.647245
```

```r
mean(backpackWT)+2*sd(bootmeans)
```

```
## [1] 16.85276
```

```r
mean(backpackWT)-2*(sd(backpackWT)/sqrt(20))
```

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

```r
mean(backpackWT)+2*(sd(backpackWT)/sqrt(20))
```

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

The interval is (9.553677, 16.94632) for the bootstrapped SE. The CI for the standard error from the formula is (9.461166, 17.03883). The bootstrapped CI is slightly tighter. This makes sense since we have more information.

Summary

The interval is (9.553677, 16.94632) for the bootstrapped SE. The CI for the standard error from the formula is (9.461166, 17.03883). This means out of an infinite number of bootstraps 95% of the means will fall within this interval.