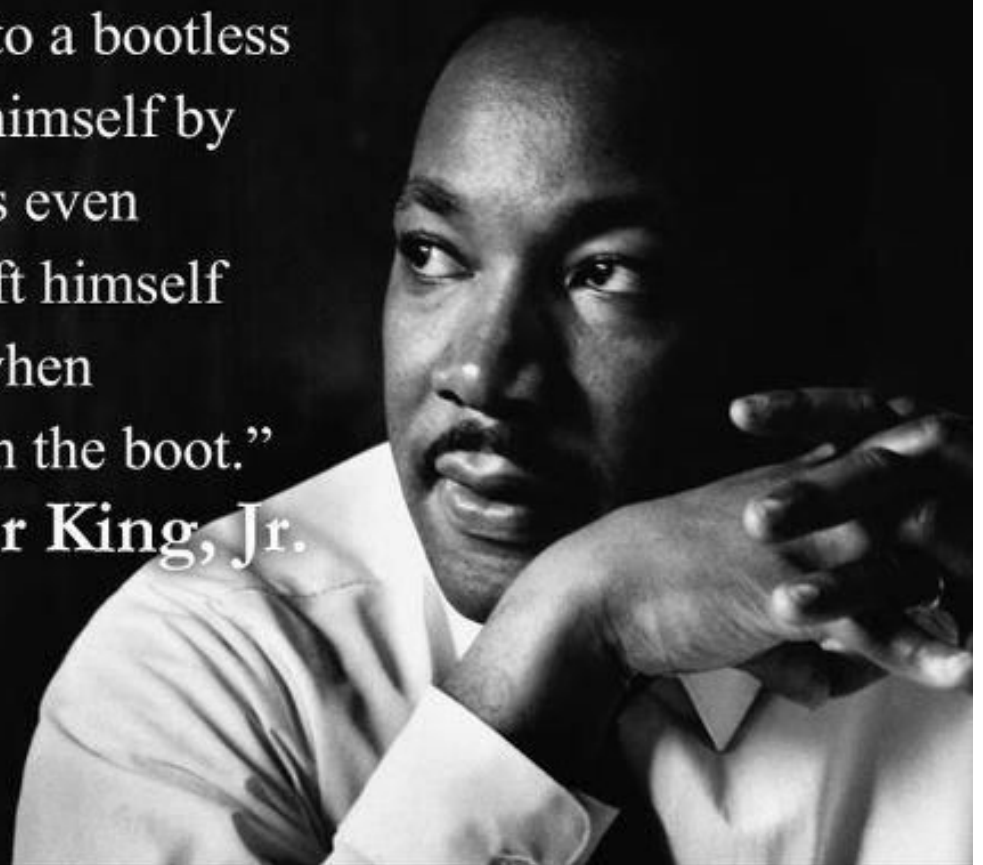


The Bootstrap

STAT315 02/04

“It is a cruel jest to say to a bootless man that he should lift himself by his own bootstraps. It is even worse to tell a man to lift himself by his own bootstraps when somebody is standing on the boot.”

-Martin Luther King, Jr.



The **bootstrap** is similar to cross-validation, but simpler. It is a technique for calculating things like confidence intervals when the assumptions on which they are based are questionable.

Example: weights of 10 halibut in kg:

1.1 1.4 1.8 3.0 59.6 3.2 1.1 1.0 1.1 1.0

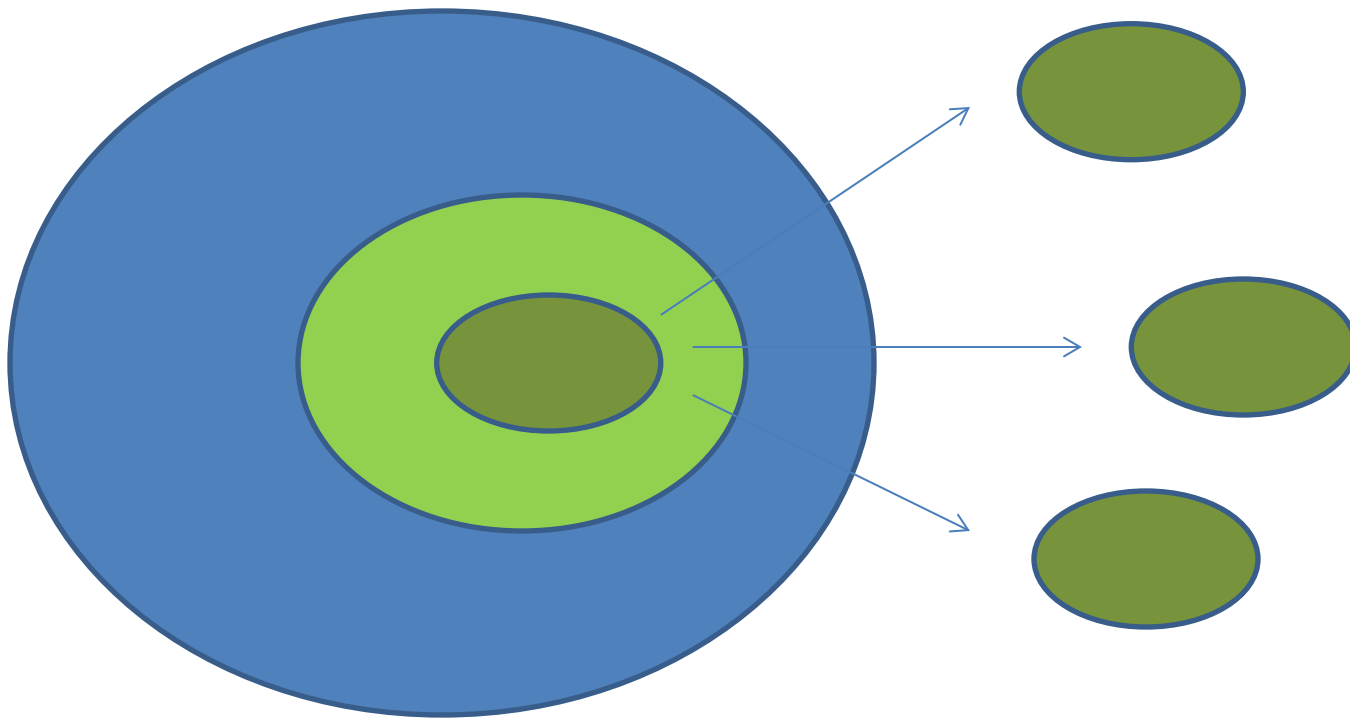
95% CI for population mean: $\bar{x} \pm t SE(\bar{x})$

$7.4 \pm 2.26(5.8) = 7.4 \pm 13$

Questions:

- What does the confidence interval mean (precisely)?
- What assumptions are being made for this calculation to be valid?
- What theorem justifies these assumptions?
- Do you think these assumptions are justified in this example?

A confidence interval is a statement about what happens when we take more samples from the same population. We can't take more samples from the same population, but remember what we did with cross-validation; we can sample from the sample, also known as **resampling**.



Since sample size is very important, it is crucial that we get samples from the sample which are of the same size as the original sample. This can be done if we sample **with replacement** from the original sample.

1.1 1.4 1.8 3.0 59.6 3.2 1.1 1.0 1.1 1.0

Bootstrap samples:

1.4 1.1 59.6 3.2 1.0 1.4 1.8 1.1 1.0 3.2

3.2 3.2 1.1 59.6 3.2 1.1 3.0 1.1 3.2 3.0

3.0 1.1 1.0 **59.6** 1.4 1.0 1.1 1.1 1.0 **59.6**

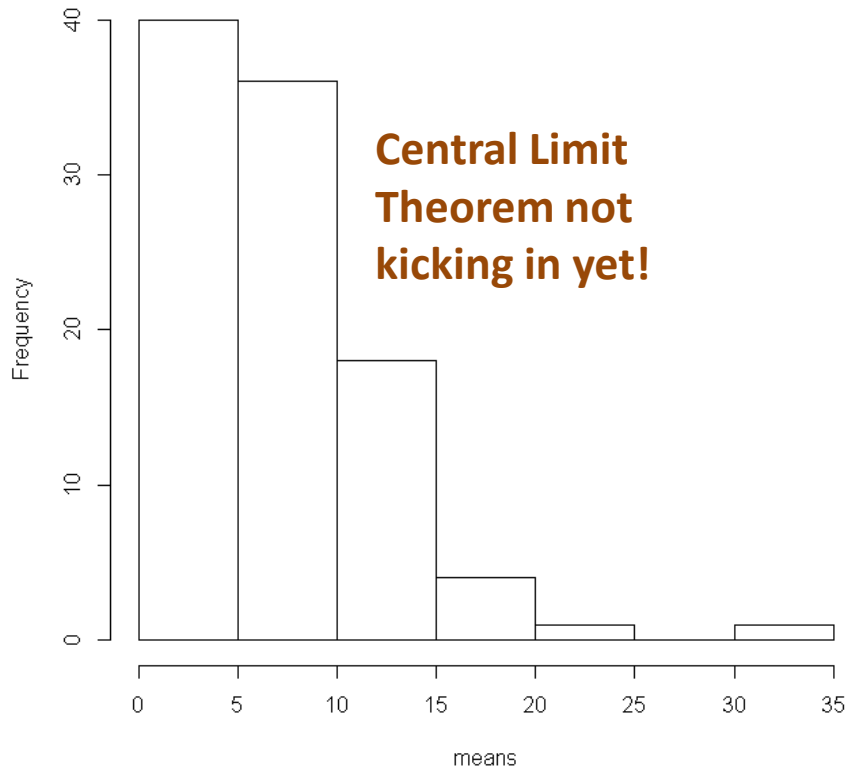
1.0 1.0 1.1 1.8 1.1 1.0 1.0 1.4 1.1 1.1

...

```
means <- rep(0,100)
for (i in 1:100)
means[i] <- mean(sample(fish,10,replace=T))

hist(means)
sort(means)[c(0.025*length(means), 0.975*length(means))]
```

Histogram of means



Bootstrap 95% CI:

(1.2, 19.1)

Skeptical question:

Given that the intervals are about the same -- $(0,20)$ versus $(1,19)$ -- doesn't that mean that they're equally good?

Answer:

Yes (or in this case equally bad; they have true coverage probability of only around 50%), but the bootstrap can be used to get an estimate for any statistic you like, and does not rely on any **distributional** assumptions. It does, however, rely on the assumption that a sample from a sample is as good as a sample from the population, which may not be the case.



Why does my bootstrap interval have terrible coverage?



I wanted to do a class demonstration where I compare a t-interval to a bootstrap interval and calculate the coverage probability of both. I wanted the data to come from a skewed distribution so I chose to generate the data as `exp(rnorm(10, 0, 2)) + 1`, a sample of size 10 from a shifted lognormal. I wrote a script to draw 1000 samples and, for each sample, calculate both a 95% t-interval and a 95% bootstrap percentile interval based on 1000 replicates.

When I run the script, both methods give very similar intervals and both have coverage probability of 50-60%. I was surprised because I thought the bootstrap interval would be better.

My question is, have I

- made a mistake in the code?
- made a mistake in calculating the intervals?
- made a mistake by expecting the bootstrap interval to have better coverage properties?

Also, is there a way to construct a more reliable CI in this situation?

```
tCI.total <- 0
bootCI.total <- 0
m <- 10 # sample size
true.mean <- exp(2) + 1

for (i in 1:1000){
```

asked 2
viewed 3

Commur
meta Why
mig

Related

- 6 Covera
basic b
Interval
- 3 p-value
obtaine
- 4 Shifting
interval
original
- 2 For boc
a high
to low
- 1 Compu
confide
predicti
report

You can use the bootstrap to work out the sampling properties of all sorts of things. Here, we are interested in using it to analyse regressions when we suspect that the regression assumptions might not be satisfied, or when we want to calculate complicated statistics.

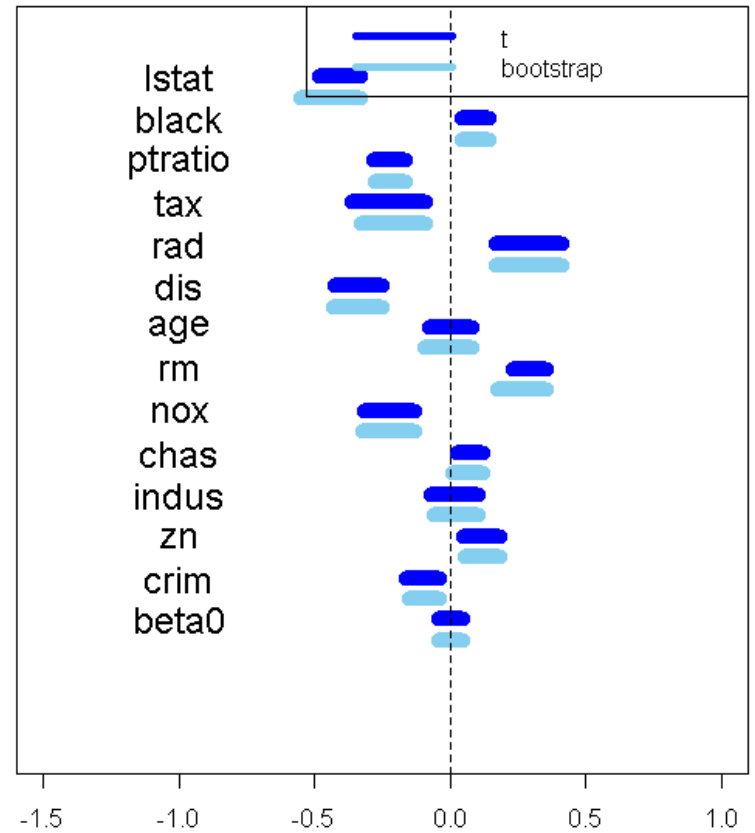
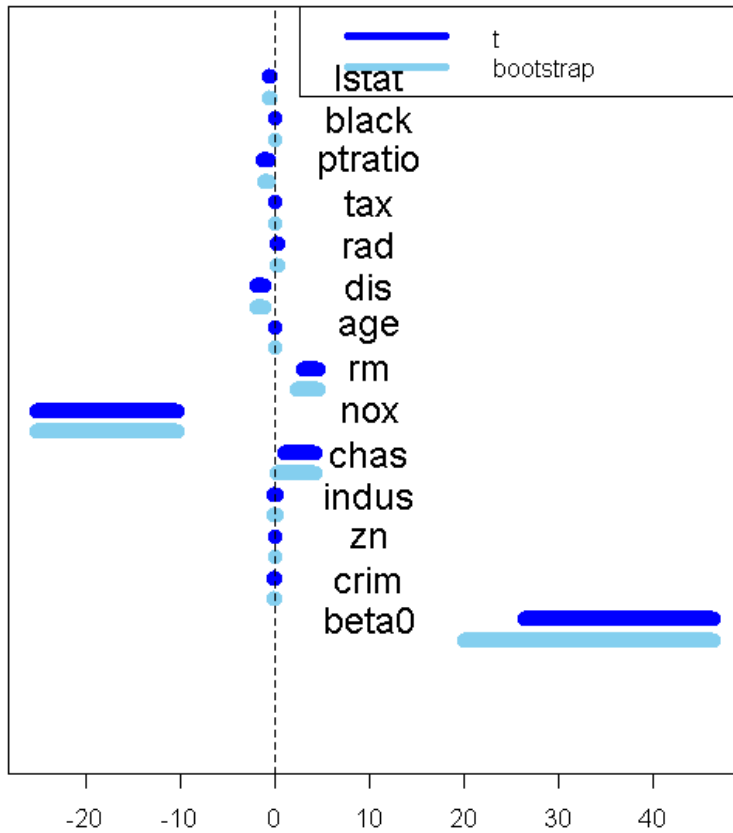
Example: Boston data, model $\text{medv} \sim \cdot$. Again.

95 % t-based CIs for all regression coefficients

```
beta0 crim  zn indus chas  nox  rm  age  dis rad  tax ptratio
[1,] 26.43 -0.17 0.02 -0.10 0.99 -25.27 2.99 -0.03 -1.87 0.18 -0.02  -1.21
[2,] 46.49 -0.04 0.07  0.14 4.38 -10.26 4.63  0.03 -1.08 0.44  0.00  -0.70
  black lstat
[1,]  0.00 -0.62
[2,]  0.01 -0.43
```

95% Bootstrap CIS

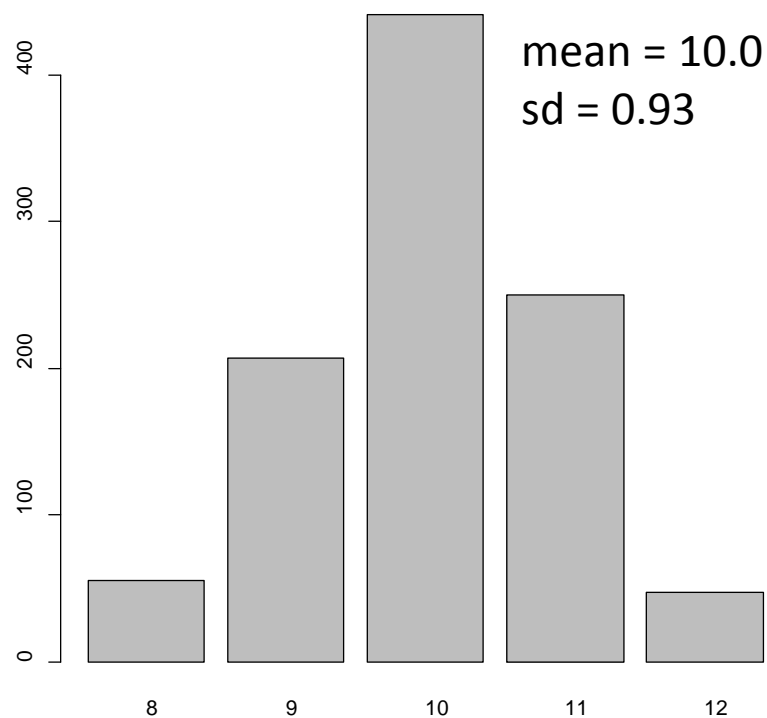
```
beta0 crim  zn indus chas  nox  rm  age  dis rad  tax ptratio
[1,] 19.38 -0.16 0.02 -0.08 0.25 -24.86 2.28 -0.03 -1.92 0.18 -0.02  -1.20
[2,] 51.58 -0.02 0.07  0.12 5.05 -10.57 5.66  0.03 -1.07 0.42 -0.01  -0.73
  black lstat
[1,]  0.00 -0.71
[2,]  0.01 -0.33
```



Left: raw data; Right: x's centered and scaled
(which is better?)

Let's suppose we are interested in some strange statistic like: how many of the regression coefficients are greater in size than $1/13$?

The bootstrap gives us an easy way to estimate this. For each bootstrap sample, we just calculate the relevant statistic.



Of course, a lot of computation is involved (*“the Bootstrap will blow a hole in your problem, provided you are willing to put up with the mess” – Efron*)