PLSC 503 – Spring 2020
Bootstrapping

January 30, 2020

The population is to the sample as the sample is to the bootstrap sample.

Practical (Nonparametric) Bootstrapping

- Draw one bootstrap sample of size N **with replacement** from the original data,
- Estimate the parameter(s) $\underset{k \times 1}{\tilde{\theta}}$,
- Repeat steps 1 and 2 R times, to get $\tilde{\theta}_r$, $r \in \{1, 2, ...R\}$, comprising elements $\tilde{\theta}_{rk}$,
- Examine the empirical characteristics of the resulting distribution(s) of $\tilde{\theta}_{rk}$.

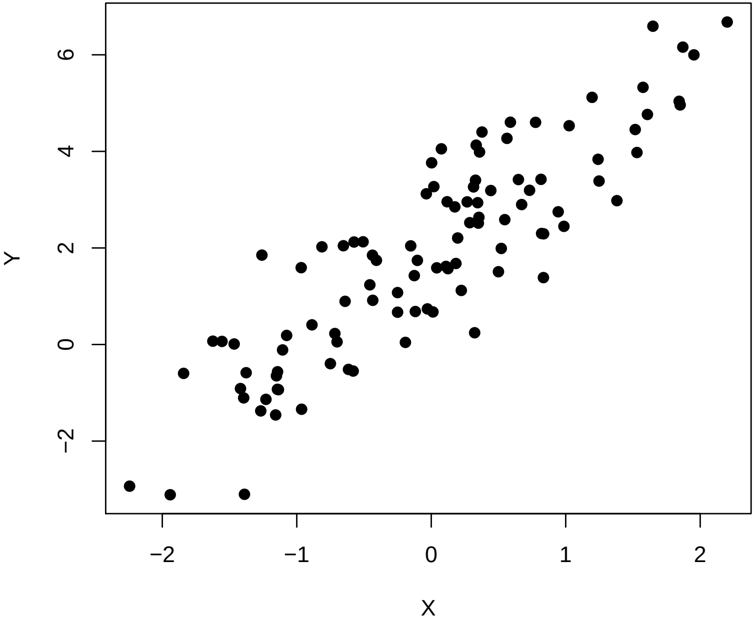
Why Bootstrap?

- **It's intuitive.**
- **It's simple.**
- **It's robust.**

Bootstrapping Simulation ($N = 100$)

```
N<-100
reps<-999

set.seed(7222009)
X<-rnorm(N)
Y<-2+2*X+rnorm(N)
```

Bootstrapping Data



Bootstrapping: "By Hand"

```
data<-data.frame(Y,X)
fitOLS<-lm(Y~X)
CI<-confint(fitOLS)

B0<-numeric(reps)
B1<-numeric(reps)

for (i in 1:reps) {
  temp<-data[sample(1:N,N,replace=TRUE),]
  temp.lm<-lm(Y~X,data=temp)
  B0[i]<-temp.lm$coefficients[1]
  B1[i]<-temp.lm$coefficients[2]
}

ByHandB0<-median(B0)
ByHandB1<-median(B1)
ByHandCI.B0<-quantile(B0,probs=c(0.025,0.975)) # <-- 95% c.i.s
ByHandCI.B1<-quantile(B1,probs=c(0.025,0.975))
```

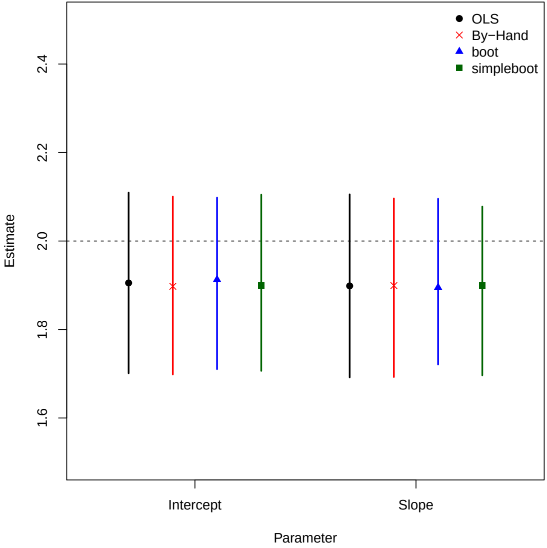
Bootstrapping Via boot

```
library(boot)

Bs<-function(formula, data, indices) { # <- regression function
    dat <- data[indices,]
    fit <- lm(formula, data=dat)
    return(coef(fit))
}

Boot.fit<-boot(data=data, statistic=Bs,
          R=reps, formula=Y~X)

BootB0<-median(Boot.fit$t[,1])
BootB1<-median(Boot.fit$t[,2])
BootCI.B0<-boot.ci(Boot.fit,type="basic",index=1)
BootCI.B1<-boot.ci(Boot.fit,type="basic",index=2)
```

Bootstrapping Via simpleboot

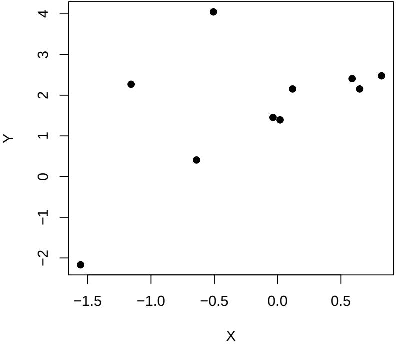
```
library(simpleboot)

Simple<-lm.boot(fitOLS,reps)
SimpleB0<-perc(Simple,.50)[1]
SimpleB1<-perc(Simple,.50)[2]
Simple.CIs<-perc(Simple,p=c(0.025,0.975))
```

Bootstrapping Results

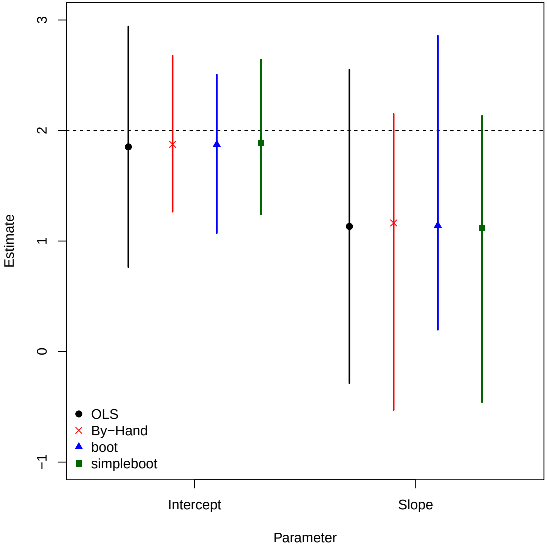


Bootstrapping: Skewed Residuals ($N = 10$)

```
> N<-10
> reps<-999
>
> set.seed(7222009)
> X<-rnorm(N)
> ustar<-rchisq(N,1) # <- skewed residuals
> Y<-2+2*X+(ustar-mean(ustar))
```



Skewed Residuals: Results



Bootstrap SCOTUS Example

```
> library(RCurl)
> temp<-getURL("https://raw.githubusercontent.com/PrisonRodeo/PLSC503-2020-git/
  master/Data/Justices.csv")
> Justices<-read.csv(text=temp, header=TRUE)
> rm(temp)
>
> summary(Justices)
       name         score              civrts          econs         Neditorials
 Rehnquist: 2   Min.   :-1.000   Min.   :19.80   Min.   :34.60   Min.   : 2.000
 Black    : 1   1st Qu.:-0.470   1st Qu.:35.90   1st Qu.:43.85   1st Qu.: 4.000
 Blackmun : 1   Median : 0.330   Median :43.70   Median :50.20   Median : 6.000
 Brennan  : 1   Mean   : 0.121   Mean   :51.42   Mean   :55.75   Mean   : 8.742
 Burger   : 1   3rd Qu.: 0.625   3rd Qu.:75.55   3rd Qu.:66.65   3rd Qu.:11.500
 Burton   : 1   Max.   : 1.000   Max.   :88.90   Max.   :81.70   Max.   :47.000
 (Other)  :24
     eratio          scoresq          lnNedit
 Min.   : 0.5000   Min.   :0.0000   Min.   :0.6931
 1st Qu.: 0.7083   1st Qu.:0.1936   1st Qu.:1.3863
 Median : 1.0000   Median :0.2500   Median :1.7918
 Mean   : 2.0242   Mean   :0.4599   Mean   :1.8441
 3rd Qu.: 2.5000   3rd Qu.:0.8281   3rd Qu.:2.4414
 Max.   :11.7500   Max.   :1.0000   Max.   :3.8501
```

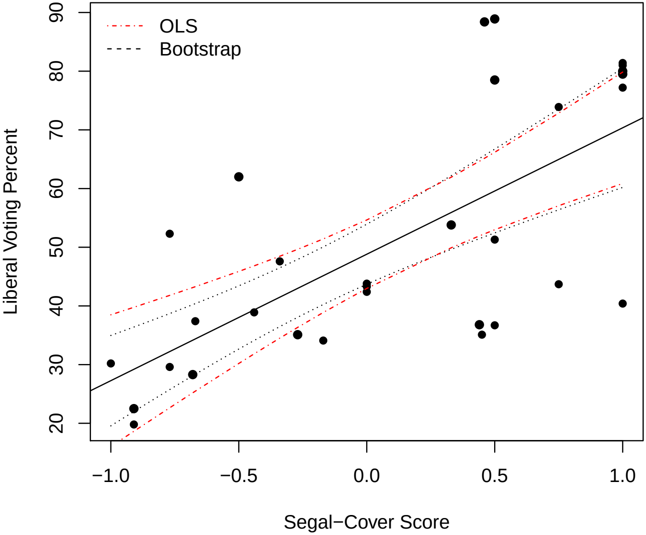
Bootstrap SCOTUS Example

```
> JOLS<-with(Justices, lm(civrts~score))
> JOLShats<-predict(JOLS,interval="confidence")
> summary(JOLS)

Call:
lm(formula = civrts ~ score)

Residuals:
    Min      1Q  Median      3Q     Max 
-29.954  -8.088  -2.120   9.396  29.680 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   48.810      2.852  17.113  < 2e-16 ***
score         21.544      4.206   5.122 1.81e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 15.63 on 29 degrees of freedom
Multiple R-squared:  0.475,Adjusted R-squared:  0.4569 
F-statistic: 26.24 on 1 and 29 DF,  p-value: 1.806e-05
```

Bootstrap SCOTUS Example

```
> JBoot <- lm.boot(JOLS, reps)
> summary(JBoot)

BOOTSTRAP OF LINEAR MODEL  (method = rows)

Original Model Fit
------------------
Call:
lm(formula = civrts ~ score)

Coefficients:
(Intercept)        score
      48.81        21.54

Bootstrap SD's:
(Intercept)        score
   2.561338     3.709290
```

Conventional and Bootstrap CIs



Bootstrapping Miscellanea

In general, use bootstraps when:

- Samples are small
- Residuals are "strange"
- Standard errors are otherwise hard to derive / estimate.

References, practical and otherwise:

- Carpenter, James, and John Bithell. 2000. "Bootstrap Confidence Intervals: When, Which, What? A Practical Guide for Medical Statisticians." *Statistics in Medicine* 19:1141-1164.
- Chernick, Michael R., and Robert A. LaBudde. 2011. *An Introduction to Bootstrap Methods with Applications to R*. New York: Wiley.
- Davison, A.C., and D.V. Hinkley. 1997. *Bootstrap Methods and their Application*. New York: Cambridge University Press.
- Efron, Bradley, and R.J. Tibshirani. 1993. *An Introduction to the Bootstrap*. London: Chapman and Hall.