PLSC 503 – Spring 2017
Bootstrapping

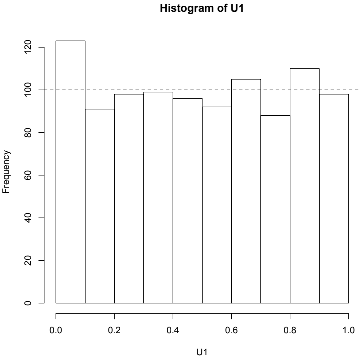
February 2, 2017

What We'll Do Today

- Review "Homework Zero"
- Bootstrapping

Generate 1000 i.i.d. $u_j \sim U(0, 1)$

```
seed<-07222009
set.seed(seed)
U1<-runif(1000)
hist(U1)
abline(h=100,lty=2)
# etc.
```



Repeat 999 more times, saving each set of draws

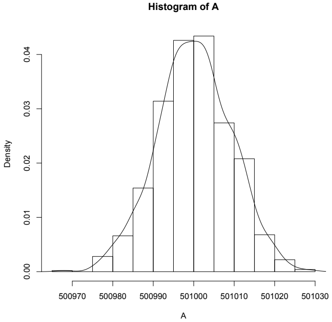
```
listU<-paste("U",1:1000,sep="")
U<-sapply(listU, function(U) U<-runif(1000))

# or

U<-data.frame(matrix(nrow=1000,ncol=1000))
colnames(U)<-paste("U",1:1000,sep="")
for (i in 1:1000) {
 U[,i]<-runif(1000)
 }
```

Create V_i by adding the integer corresponding to the order of the observation to the value of u_{ij}.

```
Seq<-seq(1,1000,1)
V<-U+Seq
```

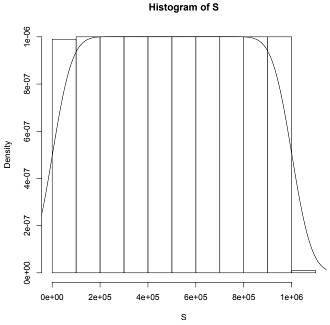
Generate an object A...where the ith entry is $A_i = \sum_{j=1}^{1000} V_{ij})$

```
A<-numeric(1000)
for(i in 1:1000) {
   A[i]<-sum(V[,i])  }
hist(A,freq=FALSE)
lines(density(A))
```



Create a second object S...where the jth entry is $S_j = \sum_{i=1}^{1000} V_{ij}$.

```
S<-numeric(1000)
for(i in 1:1000) {
   S[i]<-sum(V[i,])  }
hist(S,freq=FALSE)
lines(density(S))
```

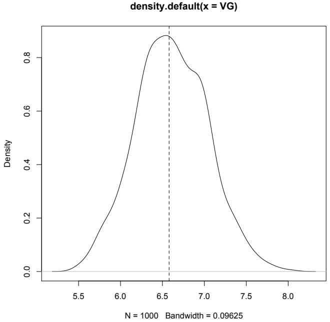


Transform your 1000 bundles of $U(0,1)$ draws into 1000 bundles G_{ij} of draws from a Gumbel(1,2) distribution.

```
G <- 1-2*(log(-log(U)))
```

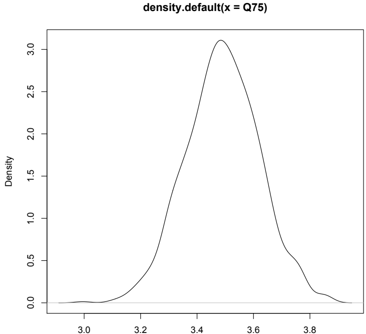
Plot the empirical variances of G...

```
VG<-numeric(1000)
for(i in 1:1000) {
  VG[i]<-var(G[,i])  }
plot(density(VG))
abline(v=((3.14159265^2) / 6) * 4,lty=2)
```



Plot the density of the values of the 75th percentiles...

```
Q75<-numeric(1000)
for(i in 1:1000) {
  Q75[i]<-quantile(G[,i], .75)
  }
plot(density(Q75))
```

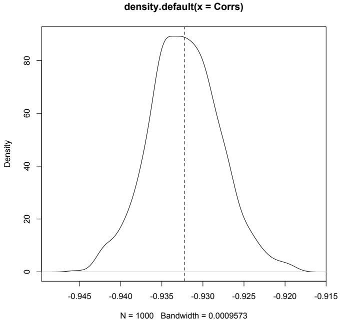


Generate 1000 draws $Y_{ij} = -2G_{ij} + \epsilon_{ij}$, $\epsilon_{ij} \sim N(0, 4)$.

```
Y<-(-2*G)+(rnorm(1000,mean=0,sd=2))
```

Plot...the distribution of the 1000 Pearson correlations between Y and G

```
Corrs<-numeric(1000)
for(i in 1:1000){
  Corrs[i]<-cor(G[,i],Y[,i])
     }
plot(density(Corrs))
abline(v=mean(Corrs),lty=2)
```



Bootstrapping

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.**

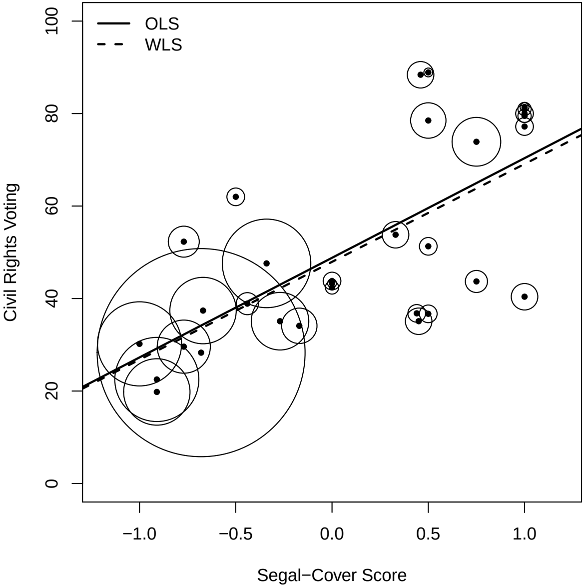
Bootstrap Example

```
> Justices<-read.csv("Justices.csv")
> attach(Justices)
> summary(Justices)
     name               score              civrts          econs      
 Length:31          Min.   :-1.0000   Min.   :19.80   Min.   :34.60  
 Class :character   1st Qu.:-0.4700   1st Qu.:35.90   1st Qu.:43.85  
 Mode  :character   Median : 0.3300   Median :43.70   Median :50.20  
                    Mean   : 0.1210   Mean   :51.42   Mean   :55.75  
                    3rd Qu.: 0.6250   3rd Qu.:75.55   3rd Qu.:66.65  
                    Max.   : 1.0000   Max.   :88.90   Max.   :81.70  
  Neditorials         eratio            scoresq           lnNedit      
 Min.   : 2.000   Min.   : 0.5000   Min.   :0.0000   Min.   :0.6931  
 1st Qu.: 4.000   1st Qu.: 0.7083   1st Qu.:0.1936   1st Qu.:1.3863  
 Median : 6.000   Median : 1.0000   Median :0.2500   Median :1.7918  
 Mean   : 8.742   Mean   : 2.0242   Mean   :0.4599   Mean   :1.8442  
 3rd Qu.:11.500   3rd Qu.: 2.5000   3rd Qu.:0.8281   3rd Qu.:2.4414  
 Max.   :47.000   Max.   :11.7500   Max.   :1.0000   Max.   :3.8501  
```

Bootstrap Example

```
> OLSfit<-with(Justices, lm(civrts~score))
> summary(OLSfit)

Call:
lm(formula = civrts ~ score)

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
```

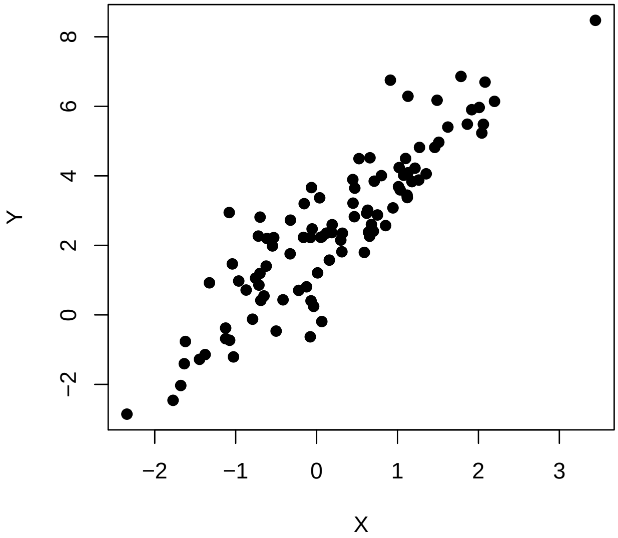
Figure: Plot of civrts Against score, Weighted by Neditorials



Bootstrapping: "By Hand"

```
N<-100
reps<-999

set.seed(1337)
X<-rnorm(N)
Y<-2+2*X+rnorm(N)
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 "By Hand"



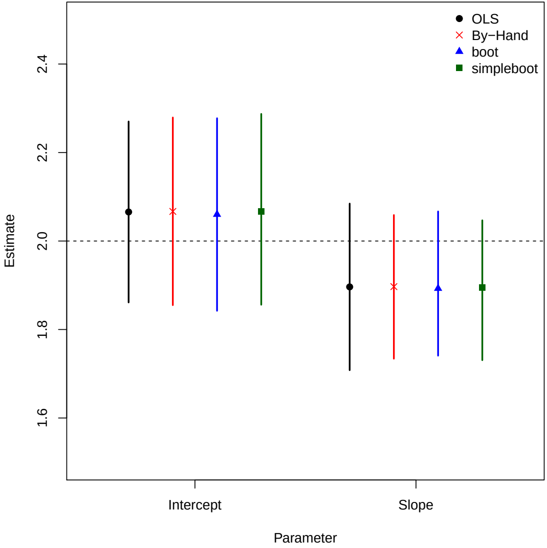
Bootstrapping Via boot

```
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`

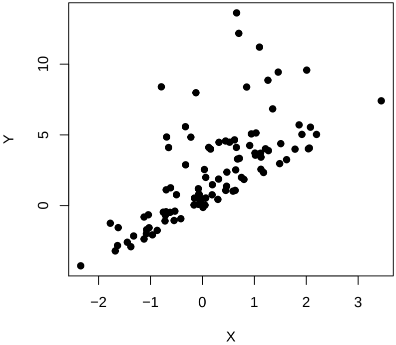
```
library(simpleboot)

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<-100
reps<-999

set.seed(1337)
X<-rnorm(N)
ustar<-rchisq(N,2) # <- skewed u.s
Y<-2+2*X+(ustar-mean(ustar))
data<-data.frame(Y,X)
fitOLS<-lm(Y~X)
CI<-confint(fitOLS)
```



Skewed Residuals: Results