HW3 STAT5376

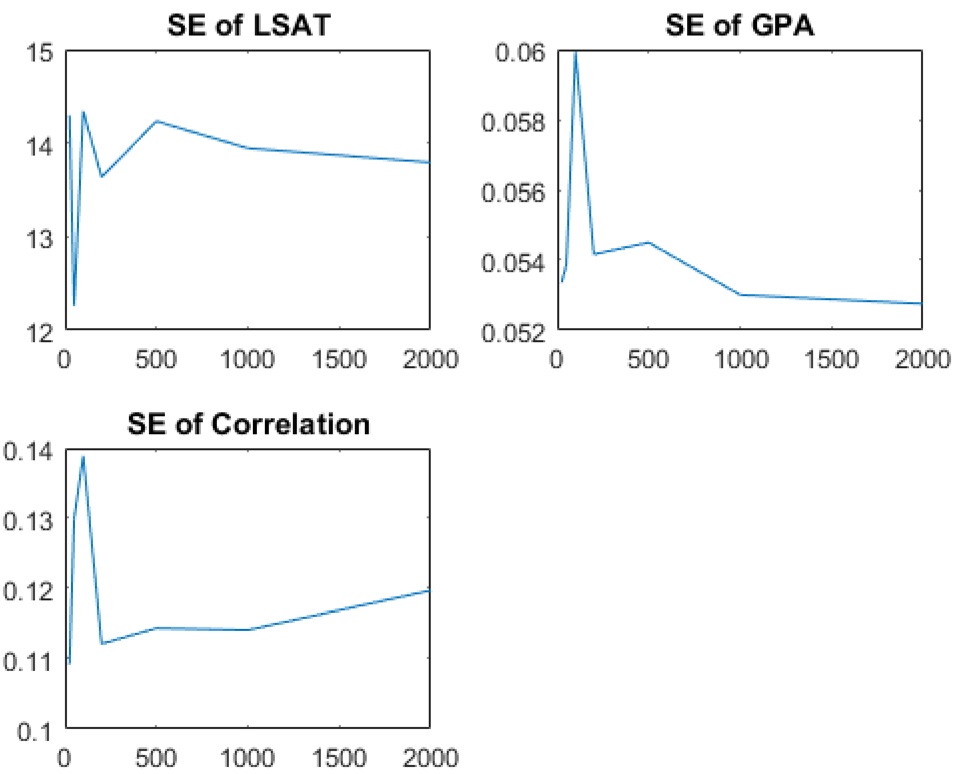
Bootstrap

Li Sun

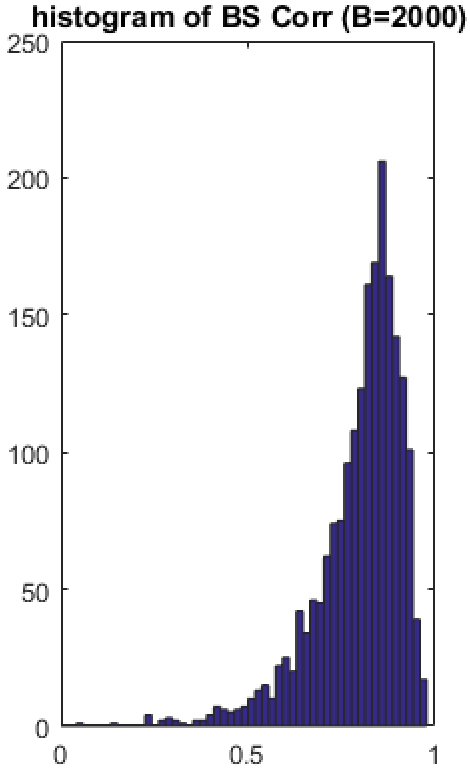
November 6, 2016

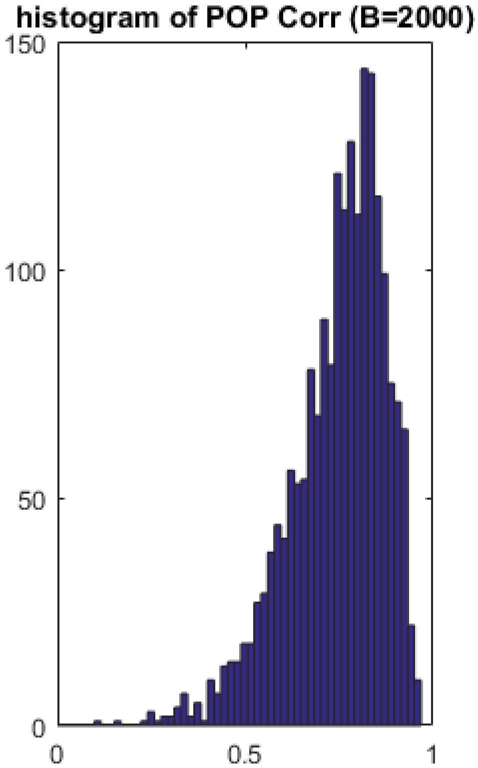
1. Consider the law school data. Bootstrap it with n=15 and compute
(a)means and correlation coefficient from the sampled data
$\mu_x = 603.8667$
$\mu_y = 3.142$
$\rho = 0.4753$

(b)SE of estimates



(c)Plot histogram of bootstrap replicates of corr for the case B=2000.





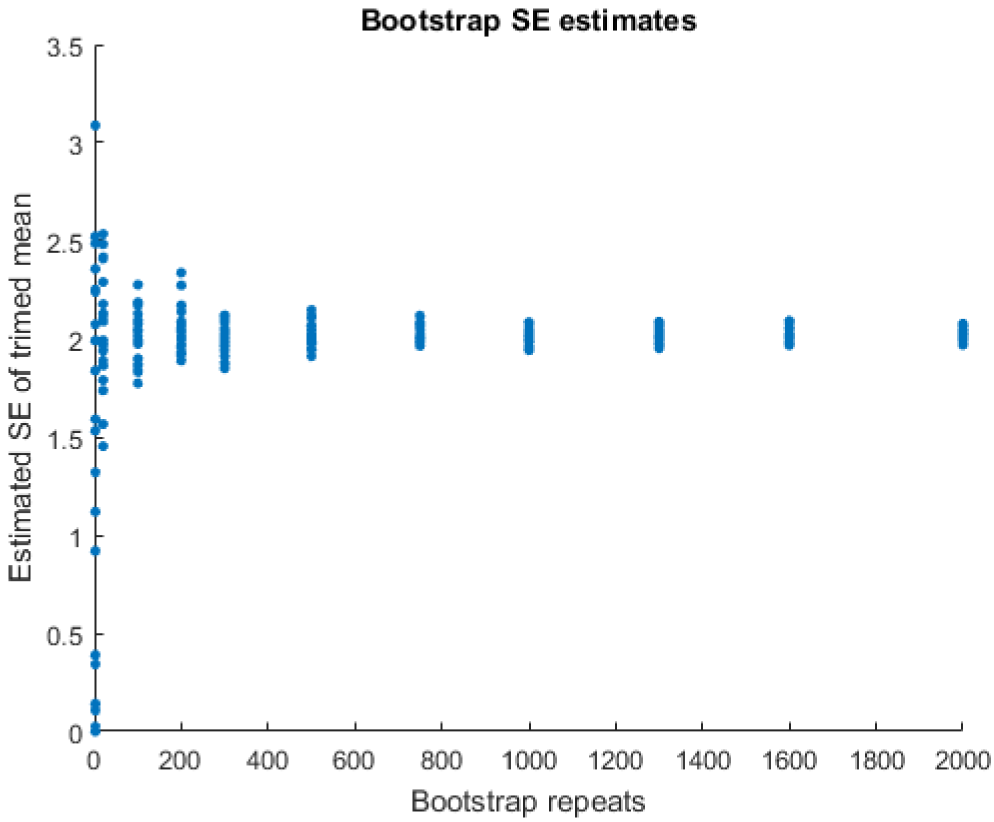
The two histogram centered at similar location but the first one(Corr estimates from Bootstrapping) seems to be spikier(smaller standard error). But generally very similar.

2. Estimating trimmed mean

(a)

B	25	100	200	500	1000	2000
SE	2.2349	2.0212	1.8934	2.0582	2.0164	2.0564

(b)How large should we take B to provide reasonable accuracy.



According to this figure, we can see the estimated of standard error of trimed mean estimator converges quickly when B goes up. And when B is above 1000, it seems good enough.

3. Estimating the bias of median estimating the mean.

I use 1000 repeats for bootstrap median estimate.
when n=10, Bias=0.0906
when n=20, Bias=0.1063
when n=100, Bias=-0.0574
The bias is based on the sample we simulated from the normal distribution.

All code please see https://github.com/rikku1983/STAT5376

Thanks!

```
clear all; close all;
load('lawschooldata.mat');
% Q1
a1=datasample(X,15);
mean(a1);
corr(a1);

B=[25 50 100 200 500 1000 2000];
for j=1:7
    for i=1:B(j)
        sample=datasample(a1,15);
        mx(i)=mean(sample(:,1));
        my(i)=mean(sample(:,2));
        rho(i)=corr(sample(:,1), sample(:,2));
    end
    mxse(j)=std(mx);
    myse(j)=std(my);
    rhose(j)=std(rho);
end
subplot(2,2,1),plot(B,mxse);title('SE of LSAT');
subplot(2,2,2),plot(B,myse);title('SE of GPA');
subplot(2,2,3),plot(B,rhose);title('SE of Correlation');

close all;
for i=1:2000
    sample=datasample(X,15);
    rhopop(i)=corr(sample(:,1), sample(:,2));
end
subplot(121),hist(rho, 60);title('histogram of BS Corr
(B=2000)');xlim([0,1]);
subplot(122),hist(rhopop, 60);title('histogram of POP Corr
(B=2000)');xlim([0,1]);
%rhotrue=corr(X(:,1),X(:,2));
%std(rho)
%std(rhopop)

%Q2
close all; clear all;
dat2=[1 2 3.5 4 7 7.3 8.6 12.4 13.8 18.1];
B=[25 100 200 500 1000 2000];
for j=1:6
    for i=1:B(j)
        sample=sort(datasample(dat2,10));
        tm(i)=mean(sample(3:8));
    end
    se(j)=std(tm);
end

close all; clear all;
dat2=[1 2 3.5 4 7 7.3 8.6 12.4 13.8 18.1];
B=[2,20,100,200,300,500,750,1000,1300,1600,2000];
for j=1:length(B)
    for r=1:20
        for i=1:B(j)
            sample=sort(datasample(dat2,10));
            tm(i,r)=mean(sample(3:8));
```

```
        end
    end
    finalse(:,j)=std(tm);
end

scatter(reshape(repmat(B, 20,1),
length(B)*20,1),reshape(finalse,length(B)*20,1),15,'filled');
title('Bootstrap SE estimates');xlabel('Bootstrap
repeats');ylabel('Estimated SE of trimed mean');

%Q3
clear all;close all;
n=10;
dat3=normrnd(0,1,n,1);
theta=mean(dat3);
%Calculate BS estimate
B=1000;
for i=1:B
    bsdat=datasample(dat3,n);
    bstheta(i)=median(bsdat);
end
bias=theta-mean(bstheta);
bias2=theta-median(dat3);
```