**Problem1:**

(a)

For the 15 sampled data:

Mean of LSAT = 597.5488

Mean of GPA = 3.1349

Correlation coefficient = 0.9138

(b)

Using Bootstrap method to estimate the standard errors of the means and correlation coefficient for B = 25,50,100,200,500,1000 and 2000:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **SE\_mean\_LSAT** | **SE\_mean\_GPA** | **SE\_Corrcoeff** |
| **B = 25** | 51.2394 | 0.2256 | 0.0515 |
| **B = 50** | 60.3614 | 0.2367 | 0.0511 |
| **B = 100** | 55.7892 | 0.2184 | 0.0517 |
| **B = 200** | 53.1571 | 0.2253 | 0.0516 |
| **B = 500** | 56.7012 | 0.2323 | 0.0499 |
| **B = 1000** | 57.6248 | 0.2344 | 0.0502 |
| **B = 2000** | 58.8276 | 0.2464 | 0.0535 |

(c)

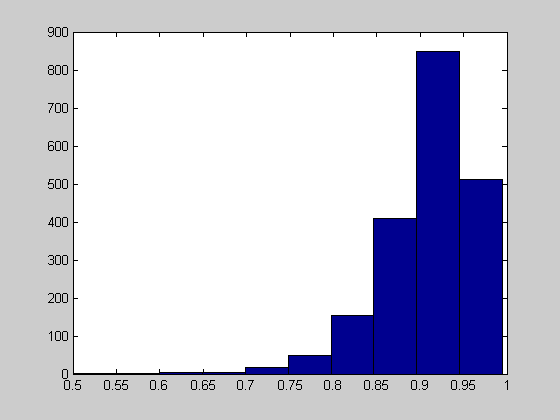


Figure 1: Histogram of bootstrap replicates of corre-coeff for B=2000

**Code for Prob1:**

clear all; close all;

load('C:\Documents and Settings\goober\Desktop\hw8\_1\_data.mat');

n=15; m=2;

%randomly select 15 sample pairs;

ind0 = floor(rand(1,15)\*length(X));

x = X(ind0(1,:),:);

% estimate the means and correlation coefficient from the sampled data;

mean\_x= mean(X);

temp = corrcoef (x(:,1), x(:,2));

cc0 = temp(1, 2);

% Bootstrap method to estimate the standard errors of these estimators

B = 25;

bx = zeros(n,m,B);

for i = 1:B

ind = randsample(n,n,'true');

bx(:,:,i) = x(ind,:);

end;

% calculate correlation coefficient

for i = 1:B

temp = corrcoef(bx(:,1,i),bx(:,2,i));

cc(i) = temp(1,2);

end;

% standard error of the correlation coefficient and means

sdcc = std(cc)

BST\_mean = mean(bx);

SE\_mean\_LSAT = std(bx(1,1,:))

SE\_mean\_GPA = std(bx(1,2,:))

% histgram of bootstrap replicates of correlation coefficients

hist(cc)

**Problem 2:**

Std\_error for the trimmed mean when B = 25, 100, 200, 500, 1000, 2000:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **B = 25** | **B = 100** | **B = 200** | **B = 500** | **B = 1000** | **B = 2000** |
| **Std\_error\_mean** | 1.0456 | 1.071 | 1.2164 | 1.2395 | 1.2208 | 1.2296 |

Repeated the bootstrapping 50 times, the variability of the estimates for the trimmed means:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **B = 25** | **B = 100** | **B = 200** | **B = 500** | **B = 1000** | **B = 2000** |
| **Variability of Std\_error** | 0.2059 | 0.0904 | 0.0662 | 0.035 | 0.039 | 0.0188 |

**Code for Prob2:**

clear all; close all;

x = [1 2 3.5 4 7 7.3 8.6 12.4 13.8 18.1];

x\_sort = sort(x);

x\_trimmed=x\_sort(3:8);

mean\_trim = mean(x\_sort(3:8));

for t=1:50

B = 2000;

bx = zeros(1,6,B);

for i=1:B

ind = randsample(6,6,'true');

bx(1,:,i) = x\_trimmed(1,ind);

end;

% trimmed mean using bootstrap;

for i=1:B

BST\_mean(i) = mean(bx(1,:,i));

end;

SE\_BST\_mean(t) = std(BST\_mean);

end;

% assess the variability of the stand eror

std(SE\_BST\_mean)

**Problem 3:**

The plots of histograms of the estimated biases for n = 10, 20 and 100:

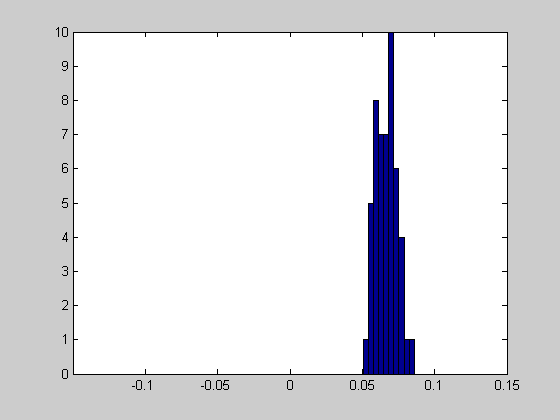


Figure 2: Histogram of the estimated bias for n = 10

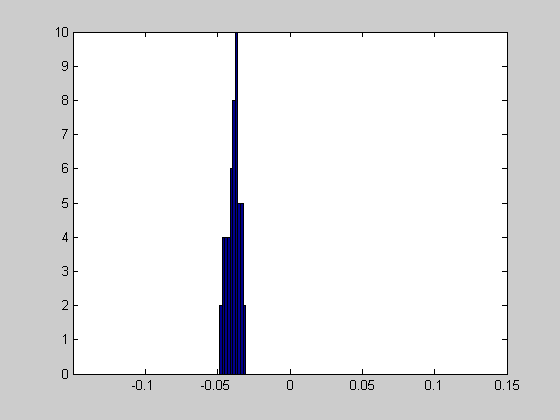


Figure 3: Histogram of the estimated bias for n = 20

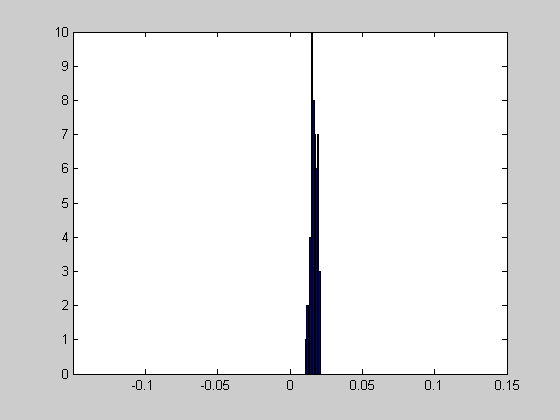


Figure 4: Histogram of the estimated bias for n = 100

**Comment:**

As the sample size increase, the biases become smaller and smaller, which can be seen from the three plots that they are more close to 0. Besides, the biases of the estimators become more concentrated, which can be explained that as the sample size increase, the bootstrap estimations become more stable.

**Code for Problem3:**

clear all; close all;

n=10;

x=randn(1,n);

theta = mean(x);

B = 2000;

for t = 1:50

bx = zeros(1,n,B);

for i = 1:B

ind = randsample(n,n,'true');

bx(1,:,i) = x(1,ind);

end

for i = 1:B

BTS\_theta(i) = median(bx(1,:,i));

end

bias(t) = mean(BTS\_theta)-theta;

end

hist(bias);

xlim([-0.15,0.15]);