Lab2

Table of Contents

Load data_lab2.mat. ACR and SAMVIQ are the averaged subjective quality scores from 28 and 39 observers respectively. A total of 192 videos were rated.

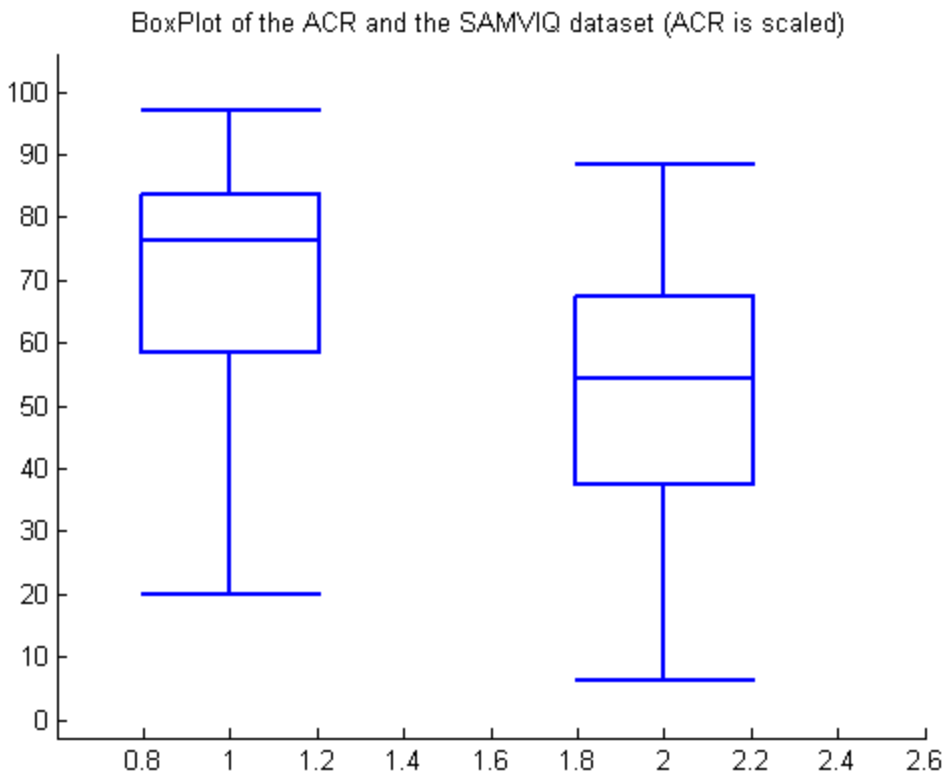
A

Compare the quality scores from ACR and SAMVIQ methodologies by creating box plots (do not use the boxplot command but instead create your own function). Which of the two exhibits higher variability around the median? Justify your answer (hint: think of the qualitative differences between the two methods, and also the fact that different number of observers were used for each method).

```
close all;
clear;

load('data_lab2.mat');

% boxplot for ACR dataset
dataset=ACR;
% sorting the data for getting the values of Q1, Q2, Q3.
sorted=sort(dataset);

len=size(sorted,1);
Q1d1=sorted(floor(len*0.25))
Q2d1=sorted(floor(len*0.50))
Q3d1=sorted(floor(len*0.75))
myboxplot1=[min(dataset) Q1d1 Q2d1 Q3d1 max(dataset)];

% boxplot for SAM dataset
dataset=SAMVIQ;
% sorting the data for getting the values of Q1, Q2, Q3.
sorted=sort(dataset);

len=size(sorted,1);
Q1d2=sorted(floor(len*0.25))
Q2d2=sorted(floor(len*0.50))
Q3d2=sorted(floor(len*0.75))
myboxplot2=[min(dataset) Q1d2 Q2d2 Q3d2 max(dataset)];

draw_data=[myboxplot1*20; myboxplot2];
draw_data=draw_data';

n = size(draw_data, 2);
```

```
unit = (1-1/(1+n))/(1+9/(4));
figure;
hold on;
for i = 1:n
    arr = draw_data(:,i);
    plot([i-unit, i+unit], [arr(5), arr(5)], 'LineWidth', 2);
    plot([i-unit, i+unit], [arr(1), arr(1)], 'LineWidth', 2);
    plot([i-unit, i+unit], [arr(3), arr(3)], 'LineWidth', 2);
    plot([i, i], [arr(5), arr(4)], 'LineWidth', 2);
    plot([i, i], [arr(2), arr(1)], 'LineWidth', 2);
    plot([i-unit, i+unit, i+unit, i-unit, i-unit], [arr(2), arr(2), arr(4), arr(4)
end
diff=max(max(draw_data))- min(min(draw_data));
%For representing the data nicely
ylim([min(min(draw_data))-(diff)/10  max(max(draw_data))+diff/10]);
title('BoxPlot of the ACR and the SAMVIQ dataset (ACR is scaled)')
figure;
```

```
Q1d1 =

    2.9259


Q2d1 =

    3.8148


Q3d1 =

    4.1852


Q1d2 =

   37.5294


Q2d2 =

   54.3333


Q3d2 =

   67.3333
```


BoxPlot of the ACR and the SAMVIQ dataset (ACR is scaled)
100
90
80
70
60
50
40
30
20
10
0
0.8
1
1.2
1.4
1.6
1.8
2
2.2
2.4
2.6

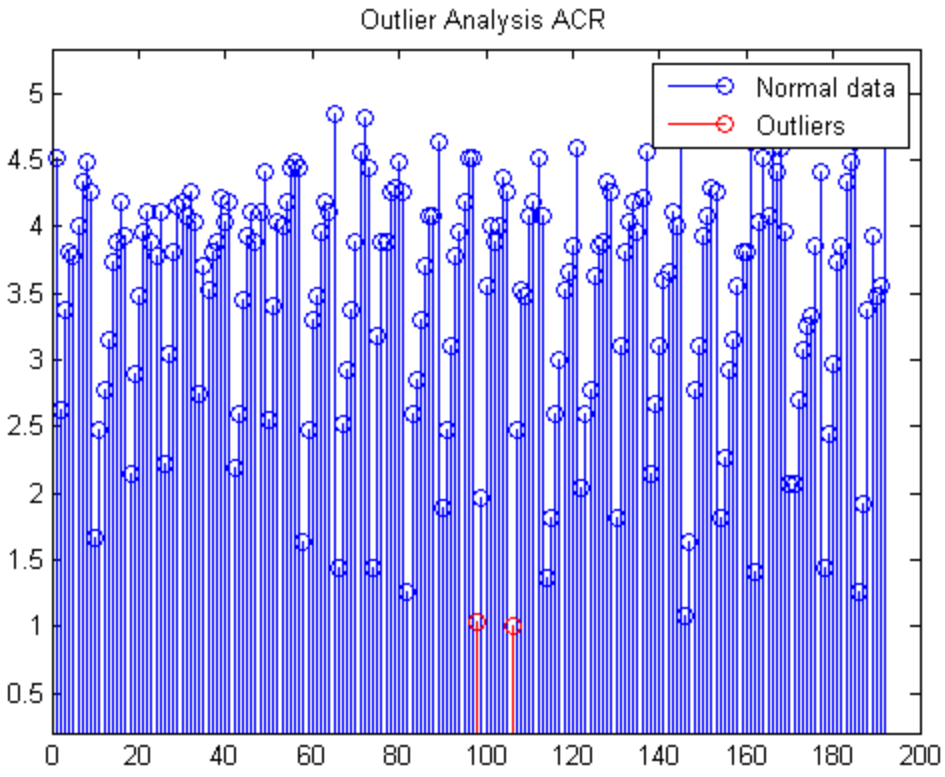
B

Which of the two, ACR or SAMVIQ, leads to mathematical outliers (assume points beyond Q3 + w*IQR or Q1 - w*IQR with w = 1.5, as mathematical outliers)? Any reason(s)?

Visualizing the outliers in the ACR dataset

```
IQR1=Q3d1-Q1d1
w=1.5;

outliers_left1= (ACR<= Q1d1 -w*IQR1);
outliers_right1=(ACR >= Q3d1 + w*IQR1);
outliers1=(outliers_left1 | outliers_right1);
outliers_data1=zeros(size(ACR));
for i=1:size(ACR,1)
    if(outliers1(i)==1)
        outliers_data1(i)=ACR(i);
    end
end
stem(ACR)
hold on
stem(outliers_data1,'ro')
legend('Normal data','Outliers')
diff=max(max(ACR)) -min(min(ACR));
%For representing the data nicely
ylim([0.2 max(max(ACR))+diff/8 ])
title('Outlier Analysis ACR')
figure
```

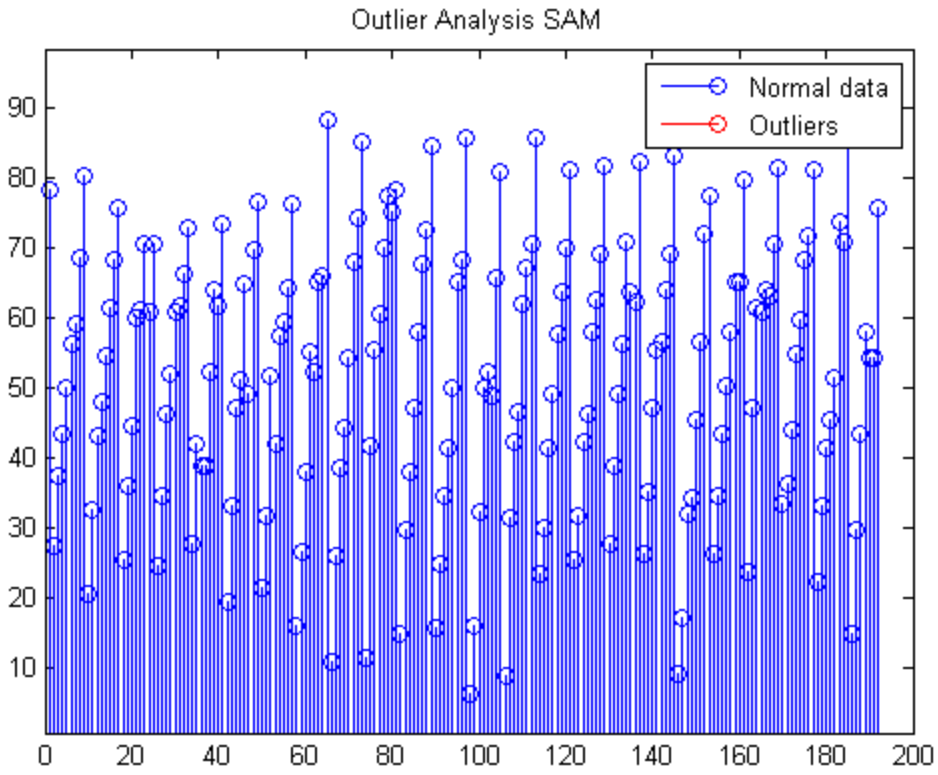
```
IQR1 =

    1.2593
```


Outlier Analysis ACR
Normal data
Outliers
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
20
40
60
80
100
120
140
160
180
200

Visualizing the outliers in the SAM dataset

```
IQR2=Q3d2-Q1d2
w=1.5;
outliers_left2= (SAMVIQ<= Q1d2 -w*IQR2);
outliers_right2=(SAMVIQ >= Q3d2 + w*IQR2);
outliers2=(outliers_left2 | outliers_right2);
outliers_data2=zeros(size(SAMVIQ));
for i=1:size(SAMVIQ,1)
    if(outliers2(i)==1)
        outliers_data2(i)=SAMVIQ(i);
    end
end
stem(SAMVIQ);
hold on;
stem(outliers_data2,'ro');
legend('Normal data','Outliers');
diff=max(max(SAMVIQ)) -min(min(SAMVIQ));
%For representing the data nicely
ylim([min(min(SAMVIQ))/10 max(max(SAMVIQ))+diff/8 ]);
title('Outlier Analysis SAM');
figure;
```

```
IQR2 =

   29.8039
```

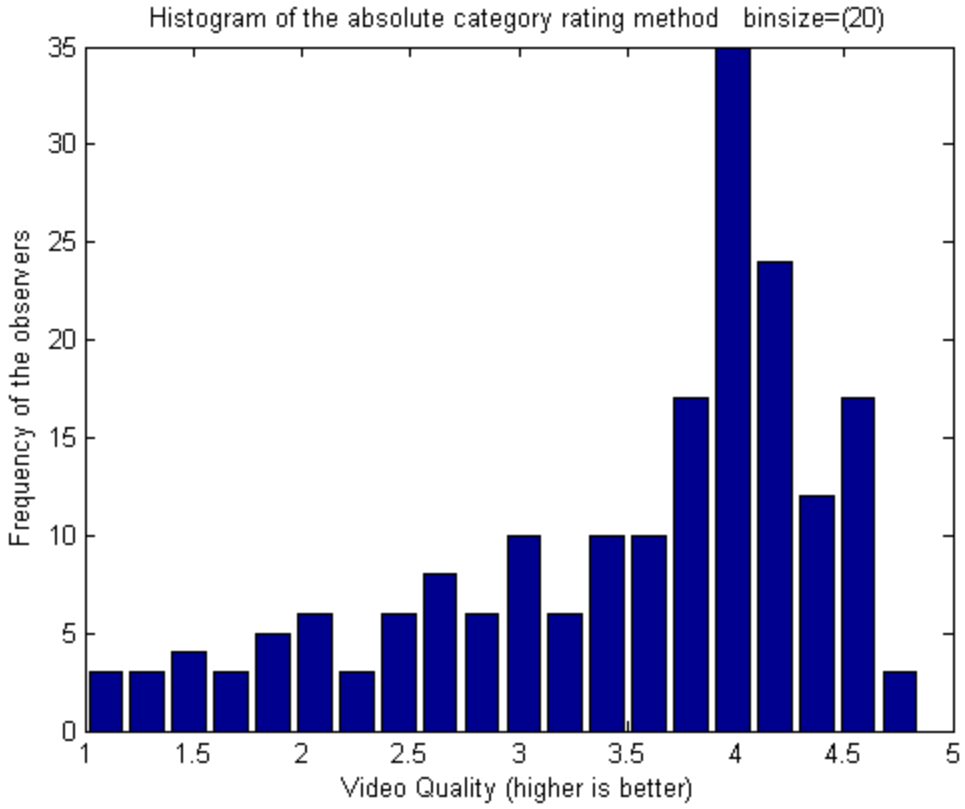

Outlier Analysis SAM
Normal data
Outliers
90
80
70
60
50
40
30
20
10
0
20
40
60
80
100
120
140
160
180
200

C

Plot the histograms for the two cases (label the axes clearly). Describe the skewness of the two histograms. Can we comment on skewness without plotting the histogram?

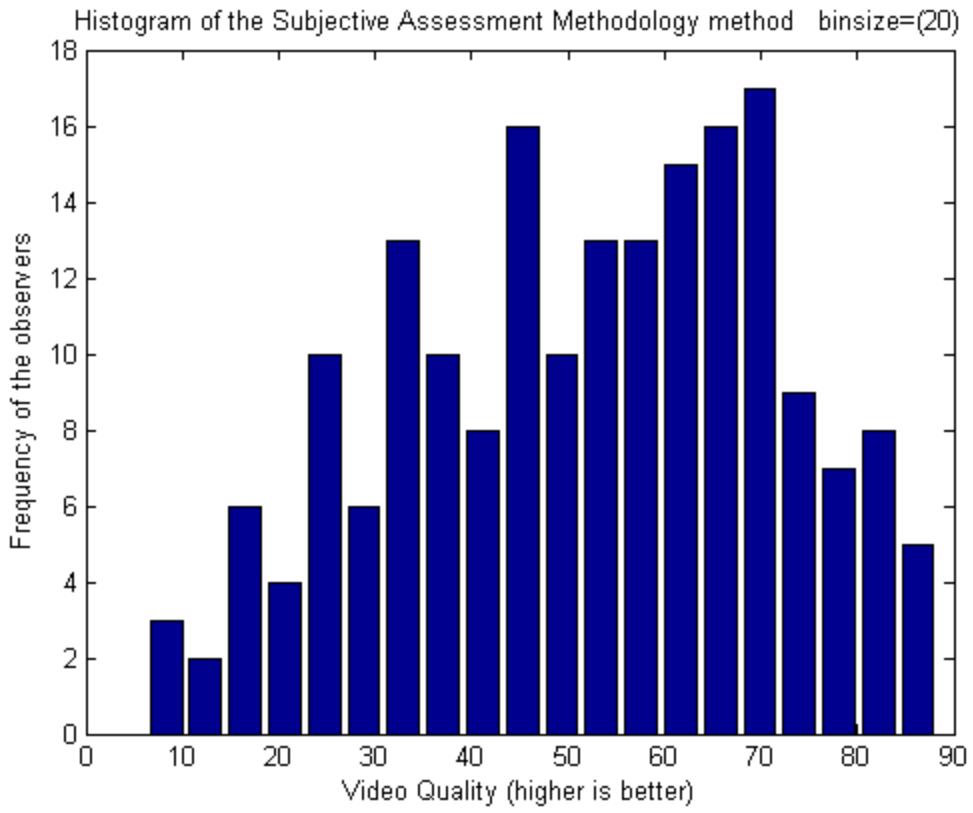
Histogram of the ACR dataset

```
data=ACR;
bins=20;
histArr=zeros(1,bins);
min1=min(min(data));
max1=max(max(data));
diff=(max1-min1)/bins;
values=min1+diff/2:diff:max1-diff/2;
for i=1:size(data,1)
    for j=1:bins
        if(data(i,1)>=min1+(j-1)*diff & data(i,1) < min1+j*diff)
            histArr(j) =histArr(j) + 1;
            break;
        end
    end
end
bar(values,histArr);
xlabel('Video Quality (higher is better)');
ylabel('Frequency of the observers');
str=strcat('Histogram of the absolute category rating method   binsize=(',int2str(
title(str);
figure;
```


Histogram of the absolute category rating method binsize=(20)
Frequency of the observers
Video Quality (higher is better)
0
5
10
15
20
25
30
35
1
1.5
2
2.5
3
3.5
4
4.5
5

Histogram of the SAM dataset

```
data=SAMVIQ;
bins=20;
histArr=zeros(1,bins);
min1=min(min(data));
max1=max(max(data));
diff=(max1-min1)/bins;
values=min1+diff/2:diff:max1-diff/2;
for i=1:size(data,1)
    for j=1:bins
        if(data(i,1)>=min1+(j-1)*diff & data(i,1) < min1+j*diff)
            histArr(j) =histArr(j) + 1;
            break;
        end
    end
end
bar(values,histArr);
xlabel('Video Quality (higher is better)');
ylabel('Frequency of the observers');
str=strcat('Histogram of the Subjective Assessment Methodology method   binsize=('
title(str);
```

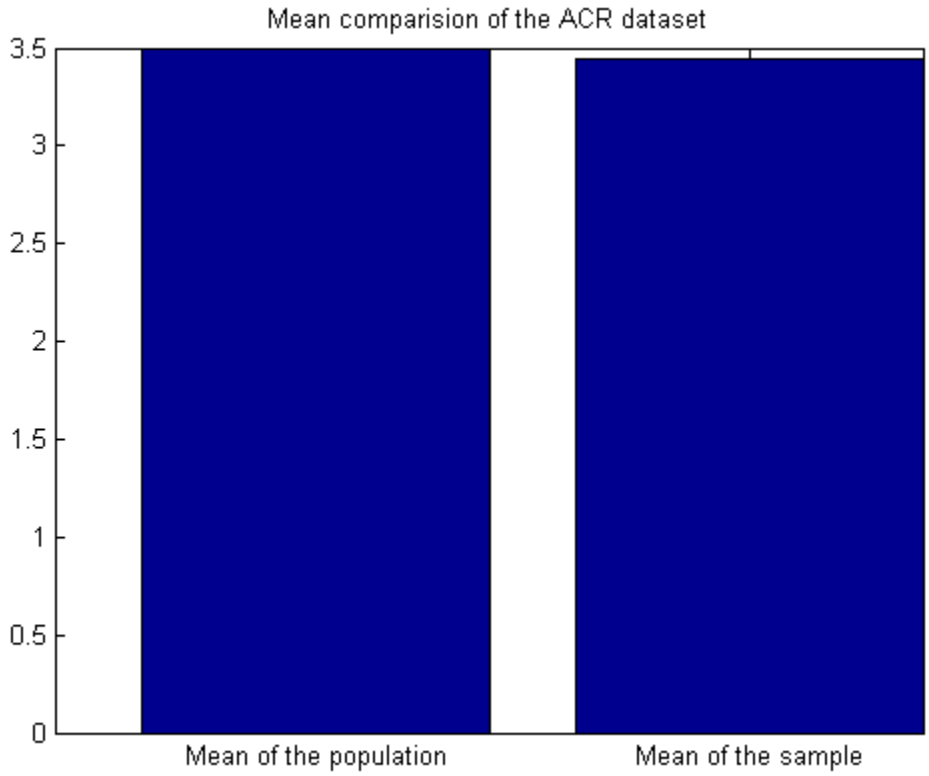


D

Randomly sample half (i.e. 96) the scores for each methodology, and compute the mean in each case. How does it compare with the mean from the original data? Based on this, can we use only 96 samples (instead of 192) to compare ACR and SAMVIQ?

Mean comparision of the ACR data

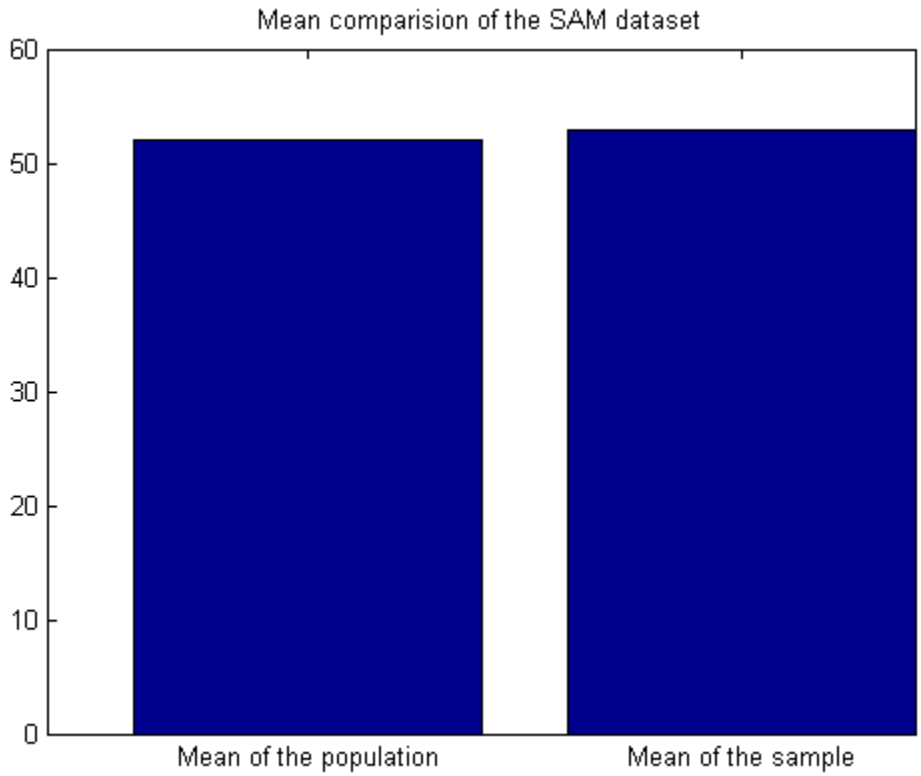
```
data=ACR;
mean_population=mean(data);
sample1=randsample(data,size(data,1)/2);
mean_sample=mean(sample1);
figure;
names={'Mean of the population' ; 'Mean of the sample'};
bar([1:2],[mean_population;mean_sample])
set(gca,'xticklabel',names)
title('Mean comparision of the ACR dataset');
```



Mean comparision of the SAM data

```
data=SAMVIQ;
mean_population=mean(data);
sample2=randsample(data,size(data,1)/2);
mean_sample=mean(sample2);
figure;
names={'Mean of the population' ; 'Mean of the sample'};
```

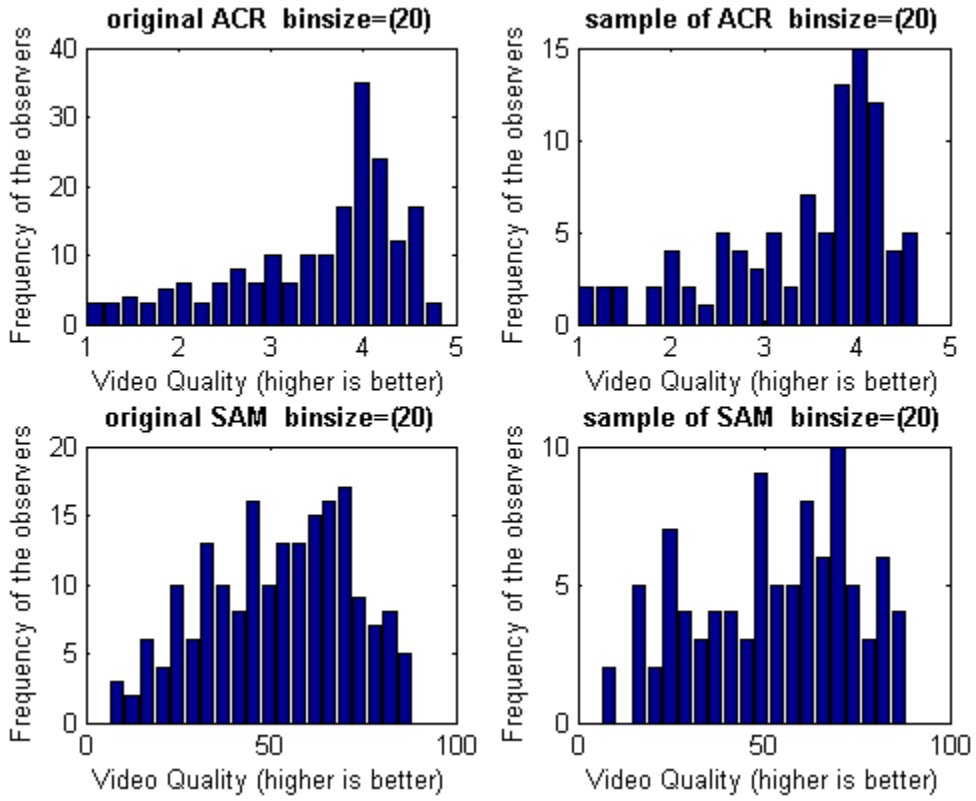
```
bar([1:2],[mean_population;mean_sample])
set(gca,'xticklabel',names)
title('Mean comparision of the SAM dataset');
```



comparing the histograms

```
subplot(2,2,1);
xaxis1='Video Quality (higher is better)';
yaxis1='Frequency of the observers';
title1=strcat('original ACR  binsize=(',int2str(bins),')');
myHist(ACR,20,xaxis1,yaxis1,title1);

subplot(2,2,2);
xaxis1='Video Quality (higher is better)';
yaxis1='Frequency of the observers';
title1=strcat('sample of ACR  binsize=(',int2str(bins),')');
myHist(sample1,20,xaxis1,yaxis1,title1);

subplot(2,2,3);
xaxis1='Video Quality (higher is better)';
yaxis1='Frequency of the observers';
title1=strcat('original SAM  binsize=(',int2str(bins),')');
myHist(SAMVIQ,20,xaxis1,yaxis1,title1);

subplot(2,2,4);
xaxis1='Video Quality (higher is better)';
yaxis1='Frequency of the observers';
```

```
title1=strcat('sample of SAM  binsize=(',int2str(bins),')');
myHist(sample2,20,xaxis1,yaxis1,title1);
```



Published with MATLAB® 7.13