BOISE STATE UNIVERSITY
DEPARTMENT OF GEOSCIENCE

Homework #1

Student name: *Hang Chen*

Course: *GEOS 422 / GEOPH 522: Data Analysis and Geostatistics*
Due date: *September 7, 2020*

Question 1

For each dataset, calculate and report the mean and standard deviation (use mean.m, std.m)

Answer. The mean of each dataset is calculated by the codes:

```
mean_depth1=mean(D1) %get the mean of depths1 and show it in command
    window
mean_depth2=mean(D2) %get the mean of depths2 and show it in command
    window
```

The results show that the mean of depths1 is 203.1770 and the mean of depths2 is 201.8116.

The standard deviation of each dataset is calculated by the codes:

```
std_depth1=std(D1) %get the standard deviation of depths1 and show it in
    command window
std_depth2=std(D2) %get the standard deviation of depths1 and show it in
    command window
```

The results show that the standard deviation of depths1 is 69.8524 and the standard deviation of depths2 is 73.0133.

Question 2

For each dataset, calculate and report the median, mode, IQR, skewness, and kurtosis.

Answer. The median of each dataset is calculated by the codes:

```
median_depth1=median(D1) %get the median of depths1 and show it in
    command window
median_depth2=median(D2) %get the median of depths2 and show it in
    command window
```

The results show that the median of depths1 is 203.4701 and the median of depths2 is 188.7049.

The mode of each dataset is calculated by the codes:

```
mode_depth1=mode(D1) %get the mode of depths1 and show it in command
    window
mode_depth2=mode(D2) %get the mode of depths2 and show it in command
    window
```

The results show that the mode of depths1 is 0.9166, and the mode of depths2 is 101.8524.

The IQR of each dataset is calculated by the codes:

```
iqr_depth1=iqr(D1) %get the IQR of depths1 and show it in command window
iqr_depth2=iqr(D2) %get the IQR of depths2 and show it in command window
```

The results show that the IQR of depths1 is 93.8922, and the IQR of depths2 is 143.0530.

The skewness of each dataset is calculated by the codes:

```
skewness_depth1=skewness(D1) %get the skewness of depths1 and show it in
    command window
skewness_depth2=skewness(D2) %get the skewness of depths2 and show it in
    command window
```

The results show that the skewness of depths1 is 0.0735, and the skewness of depths2 is 0.0535.

The kurtosis of each dataset is calculated by the codes:

```
kurtosis_depth1=kurtosis(D1) %get the skewness of depths1 and show it in
    command window
kurtosis_depth2=kurtosis(D2) %get the skewness of depths2 and show it in
    command window
```

The results show that the kurtosis of depths1 is 2.9421, and the kurtosis of depths2 is 1.1357.

Question 3

For each dataset, use the boxplot.m function to create a notched boxplot.

Answer. Notched boxplot for the depths1 by the codes:

```
figure;
boxplot(D1,'Notch','on','Labels','Depths1')% boxplot for depths1
axis([0 2 0 500]) %set axis limits
ylabel('Data value') % for the label of y axis
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
print('BOXPlot1','-dpng')
```

Notched boxplot for the depths2 by the codes

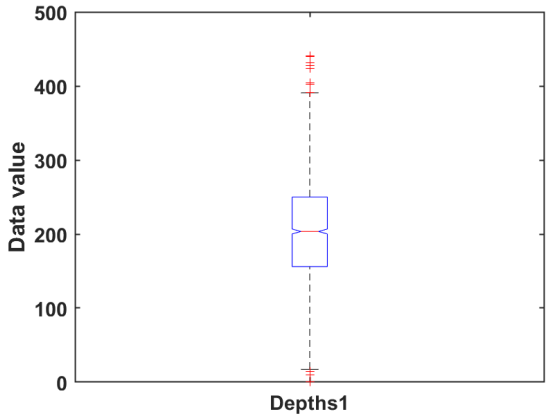

Figure 1: Boxplot for dataset depths1

```
figure;
boxplot(D2,'Notch','on','Labels','Depths2')% boxplot for depths2
axis([0 2 0 500]) %set axis limits
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
print('BOXPlot2','-dpng')
```

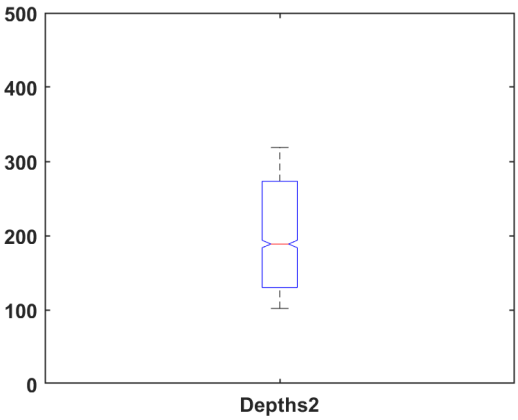

Figure 2: Boxplot for dataset depths2

Question 4

Using the hist.m function, create a histogram of each dataset, using 30 bins.

Answer. Histogram for the depths1 by the codes:

```
figure;
hist(D1,30)% histgram for depths1
xlabel('Data value')% for the label of x axis
ylabel('Counts')% for the label of y axis
axis([0 400 0 300]) %set axis limits
```

```
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
print('HISTPlot1','-dpng')
```

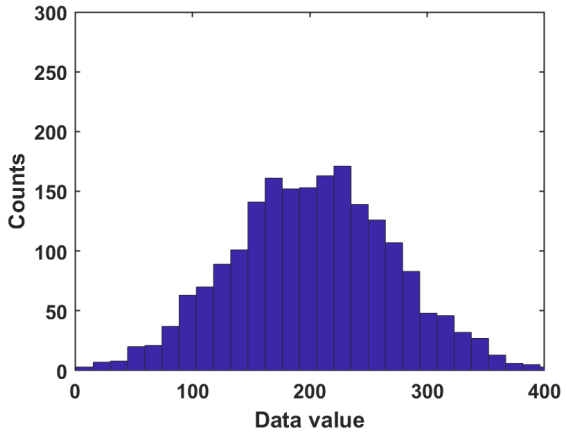

Figure 3: Histogram for dataset depths1

Histgram for the depths2 by the codes

```
figure;
hist(D2,30)% histgram for depths2
axis([0 400 0 300]) %set axis limits
xlabel('Data value')% for the label of x axis
ylabel('Counts')% for the label of y axis
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
print('HISTPlot2','-dpng')
```

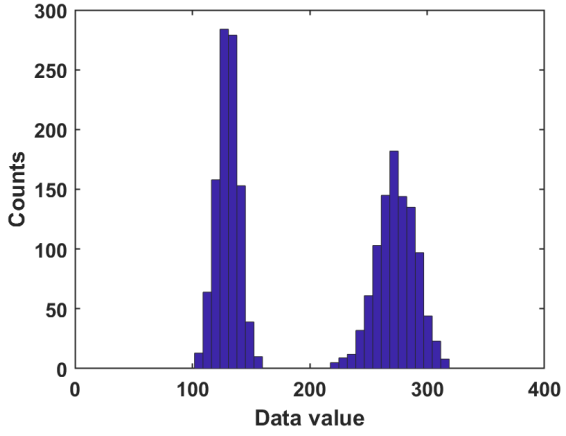

Figure 4: Histogram for dataset depths2

Question 5

Using the subplot.m function, create one figure showing all 4 plots from above.

Answer. The codes for creating one figure including 4 subfigures are

```
figure
subplot(2,2,1)
boxplot(D1,'Notch','on','Labels','Depths1')% boxplot for depths1
axis([0 2 0 500]) %set axis limits
ylabel('Data value') % for the label of y axis
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
subplot(2,2,2)
boxplot(D2,'Notch','on','Labels','Depths2')% boxplot for depths2
axis([0 2 0 500]) %set axis limits
ylabel('Data value') % for the label of y axis
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
subplot(2,2,3)
hist(D1,30)% histgram for depths1
xlabel('Data value')% for the label of x axis
ylabel('Counts')% for the label of y axis
axis([0 400 0 300]) %set axis limits
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
subplot(2,2,4)
hist(D2,30)% histgram for depths2
axis([0 400 0 300]) %set axis limits
xlabel('Data value')% for the label of x axis
ylabel('Counts')% for the label of y axis
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
print('BOXHISTPlot','-dpng')
```

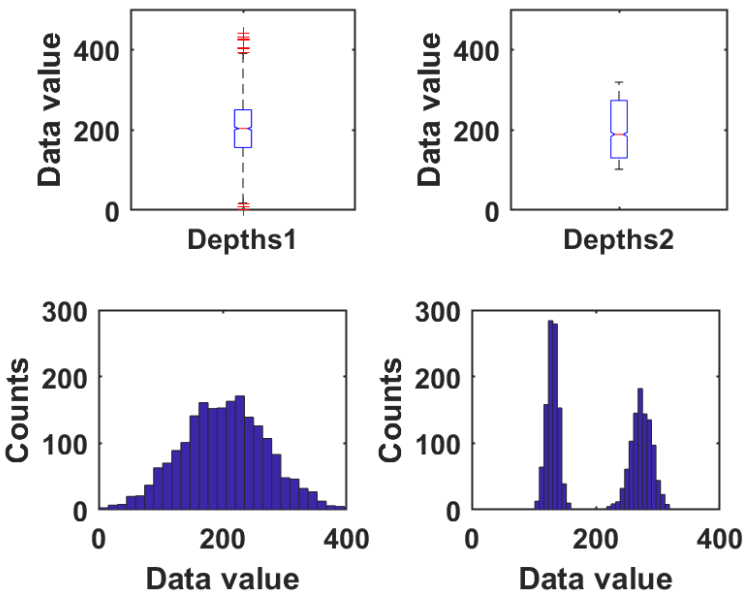

Figure 5: Box plot and Histogram for dataset depths1 and depths2

Question 6

Use the axis.m function to set the axis limits to the same range of values for all 4 plots.

Answer. I have done thie in the question 5, so the answer is same as the question 5. **Here I have a question**.

If we set the limits to the same range of values for all 4 plots. The box plot will be hard to see (e.g. figure 6). Therefore, I think the better way is to set the same range of values for box plot and set the same range values for histogram (e.g. figure 7).

```
figure
subplot(2,2,1)
boxplot(D1,'Notch','on','Labels','Depths1')% boxplot for depths1
axis([0 2 0 500]) %set axis limits
ylabel('Data value') % for the label of y axis
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
subplot(2,2,2)
boxplot(D2,'Notch','on','Labels','Depths2')% boxplot for depths2
axis([0 2 0 500]) %set axis limits
ylabel('Data value') % for the label of y axis
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
subplot(2,2,3)
hist(D1,30)% histgram for depths1
xlabel('Data value')% for the label of x axis
ylabel('Counts')% for the label of y axis
axis([0 400 0 300]) %set axis limits
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
subplot(2,2,4)
hist(D2,30)% histgram for depths2
axis([0 400 0 300]) %set axis limits
xlabel('Data value')% for the label of x axis
ylabel('Counts')% for the label of y axis
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
print('BOXHISTPlot','-dpng')
```

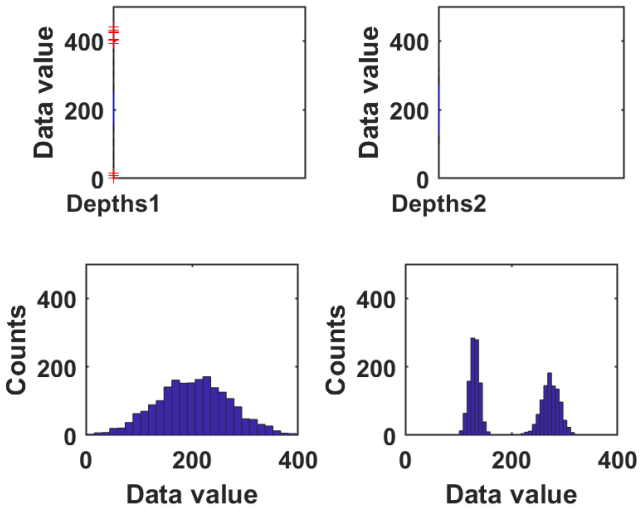

Figure 6: Box plot and Histogram for dataset depths1 and depths2

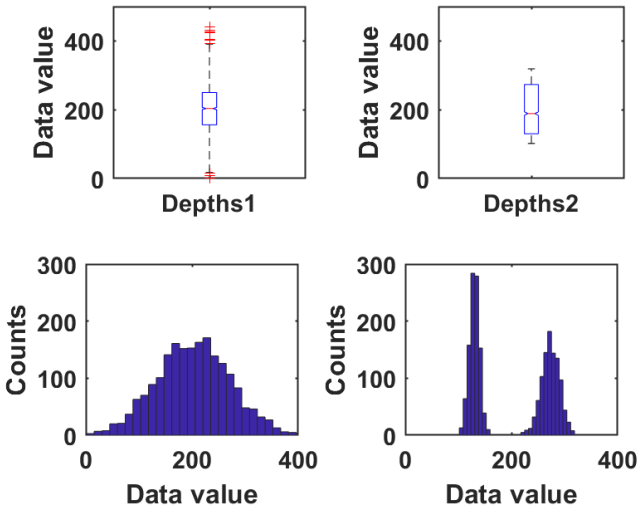

Figure 7: Box plot and Histogram for dataset depths1 and depths2

Question 7

A more useful histogram is a relative density histogram, which is normalized such that the area of the histogram is equal to 1. Using the hist function, calculate the histogram, normalize it, and plot the relative density histogram for each dataset using the bar function.

Answer. The codes for ploting the relative density histogram of two datasets

```
figure
subplot(1,2,1)
[counts,centers] = hist(D1,30); %get the counts and centers of histogram
    of depths1
binWidth = centers(2)-centers(1);% calculate the width of each bin
bar(centers,counts/2000/binWidth) %plot the relative density histogram
    of depths1
hold on
axis([0 400 0 0.02]) %set axis limits
xlabel('Data value')% for the label of x axis
ylabel('PDF')% for the label of y axis
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
subplot(1,2,2)
[counts,centers] = hist(D2,30); %get the counts and centers of histogram
    of depths2
binWidth = centers(2)-centers(1);% calculate the width of each bin
bar(centers,counts/2000/binWidth) %plot the relative density histogram
    of depths2
axis([0 400 0 0.02]) %set axis limits
xlabel('Data value')% for the label of x axis
ylabel('PDF')% for the label of y axis
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
print('RDPlot','-dpng')
```

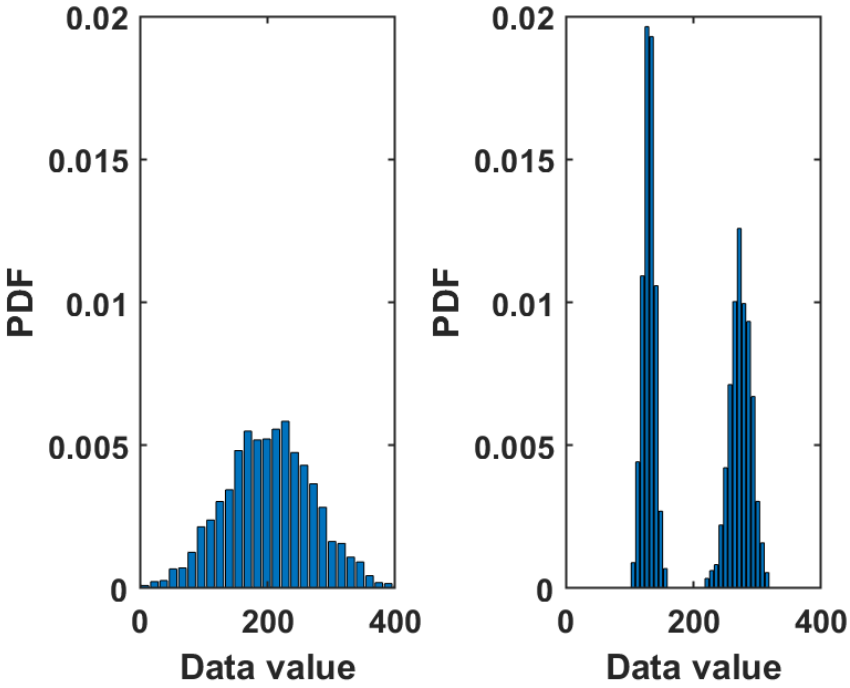

Figure 8: Box plot and Histogram for dataset depths1 and depths2

Question 8

The Gaussian (normal) distribution has the form

$$f\left(x;\mu,\sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

for mean μ and standard deviation σ. Using your calculated mean and standard deviation from above, calculate the normal distribution curve and plot it with a thick red line on your relative density histograms above.

Answer. The codes for ploting the relative density histogram of two datasets and plotting Gaussian distribution with the relative density histogram:

```
figure
subplot(1,2,1)
[counts,centers] = hist(D1,30); %get the counts and centers of histogram
    of depths1
binWidth = centers(2)-centers(1);% calculate the width of each bin
bar(centers,counts/2000/binWidth) %plot the relative density histogram
    of depths1
hold on

% plot Gaussian distribution with the relative density histogram in red
plot([0:400],mynormpdf([0:400],mean_depth1,std_depth1),'LineWidth',1,'Color','r')%
    line for dataset1
axis([0 400 0 0.02]) %set axis limits
xlabel('Data value')% for the label of x axis
ylabel('PDF')% for the label of y axis
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
subplot(1,2,2)
[counts,centers] = hist(D2,30); %get the counts and centers of histogram
    of depths2
binWidth = centers(2)-centers(1);% calculate the width of each bin
bar(centers,counts/2000/binWidth) %plot the relative density histogram
    of depths2
hold on

% plot Gaussian distribution with the relative density histogram in red
plot([0:400],mynormpdf([0:400],mean_depth2,std_depth2),'LineWidth',1,'Color','r')%
    line for dataset2
axis([0 400 0 0.02]) %set axis limits
xlabel('Data value')% for the label of x axis
ylabel('PDF')% for the label of y axis
set(gca,'LineWidth',1,'FontSize',14,'FontWeight','bold')
print('PDFPlot','-dpng')
```

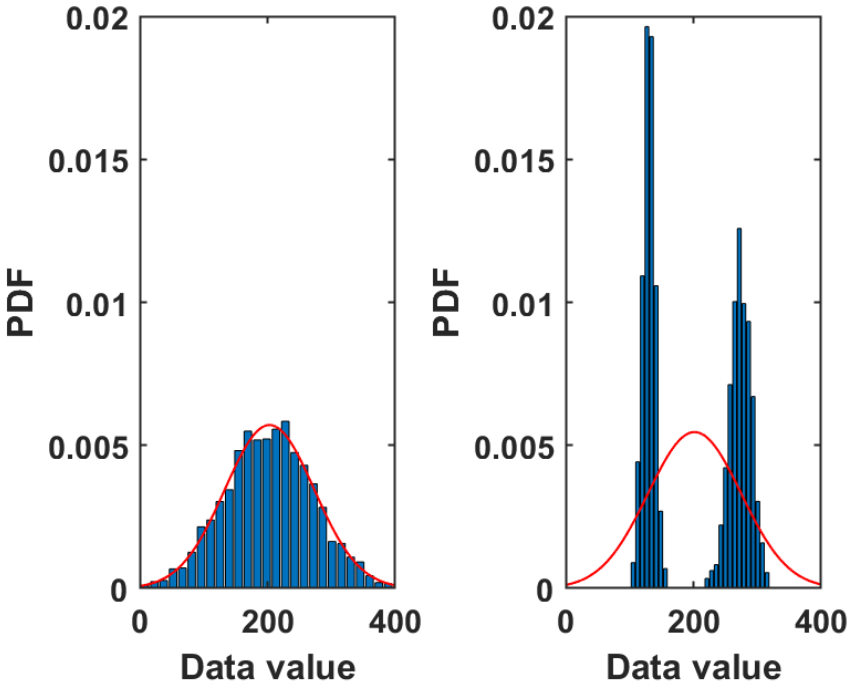

Figure 9: Box plot and Histogram for dataset depths1 and depths2

Question 9

What is the probability of a new measurement at each site being within 20cm of the average value?

Answer. The codes for the probability of a new measurement at each site being within 20cm of the average value:

```
%for depths1
pro_1_within20=length(D1(D1<(mean_depth1+20)&D1>(mean_depth1-20)))/2000
   %get the probability
%for depths2
pro_2_within20=length(D2(D2<(mean_depth2+20)&D2>(mean_depth2-20)))/2000
   %get the probability
```

The results show that the probability of a new measurement at depths1 being within 20cm of the average value is 0.2130, and the probability of that at depths2 is 1e-3.

Question 10

What is the probability of a new measurement at each site being at least 20cm larger than the average value?

Answer. The codes for the probability of a new measurement at each site being at least 20cm larger than the average value:

```
%for depths1
pro_1_out20=length(D1(D1>=(mean_depth1+20)))/2000 %get the probability
%for depths2
pro_2_out20=length(D2(D2>=(mean_depth2+20)))/2000 %get the probability
```

The results show that the probability of a new measurement at depths1 being at least 20cm larger than the average value is 0.3935, and the probability of that at depths2 is 0.4990.

Question 11

What is the probability of a new measurement at each site being at least 20cm smaller than the average value?

Answer. The codes for the probability of a new measurement at each site being at least 20cm smaller than the average value:

```
%for depths1
pro_1_smaller20=length(D1(D1<=(mean_depth1-20)))/2000 %get the
   probability
%for depths2
```

```
4 pro_2_smaller20=length(D2(D2<=(mean_depth2-20)))/2000 %get the
    probability
```

The results show that the probability of a new measurement at depths1 being at least 20cm smaller than the average value is 0.3935, and the probability of that at depths2 is 0.5000.