Probability and Statistics Homework 1

March 18, 2023

1 Question 1.a

This question is done by using MATLAB. To accomplish that, first given variables are initiated:

```
clc;
clear all;
close all;

numOfSales = [[0,2]; [2,4]; [4,6]; [6,8]; [8,10]];
number_of_real_estate_brokers = [16; 25; 13; 4; 2];
```

It is clearly seen that frequencies are equal to number of real estate brokers, and class boundaries are equal to number of sales:

```
%% A

% frequencies

frequencies = number_of_real_estate_brokers;

% class boundaries

classBoundaries = numOfSales;
```

However, class boundaries might be calculated by taking the mean of upper and lower class limits of neighbor classes. This can be useful, especially in case of that varying class intervals or multiple datasets exists. The calculation in MATLAB:

```
intervalBetweenBoundaries =  numOfSales(2,1) - numOfSales(1,2);
intervalBetweenBoundaries = intervalBetweenBoundaries/2;
intervalCount = length(numOfSales(:,1));
classBoundaries = zeros(intervalCount, 2);
for i=1:intervalCount
    classBoundaries(i,1) = numOfSales(i, 1) - intervalBetweenBoundaries;
    classBoundaries(i,2) = numOfSales(i, 2) + intervalBetweenBoundaries;
end

table(numOfSales, frequencies, classBoundaries)
```

numOfSales		frequencies	classBoundaries	
0	2	16	0	2
2	4	25	2	4
4	6	13	4	6
6	8	4	6	8
8	10	2	8	10

Class marks are simply the value in between class limits:

```
% class marks

disp(intervalCount)
classMarks = zeros(intervalCount, 1);
for i=1:intervalCount
    classMarks(i) = abs(numOfSales(i,1) + numOfSales(i,2))/2;
end
table(numOfSales, frequencies, classMarks)
```

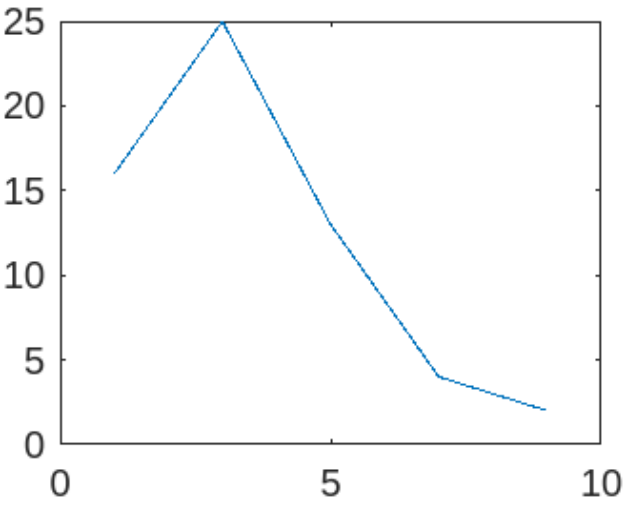
numOfSales		frequencies	classMarks
0	2	16	1
2	4	25	3
4	6	13	5
6	8	4	7
8	10	2	9

Here are class midpoints are same as class marks.

```
% class mids

classMids = classMarks;
```

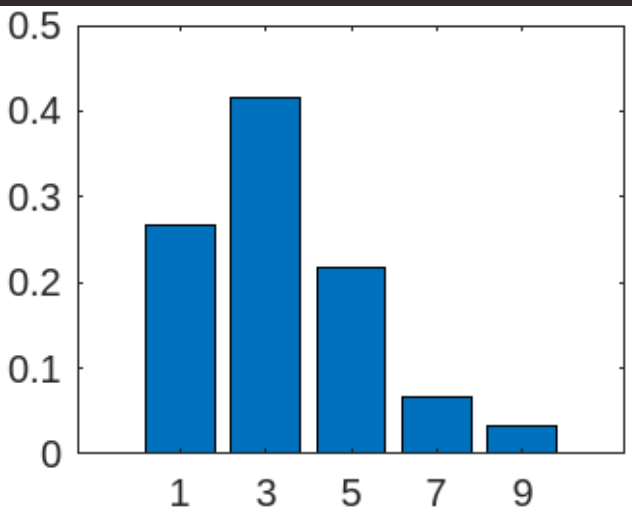
Frequency distribution can be plotted by using class marks and frequencies.

```
% frequency distribution

figure(1);
plot(classMarks, frequencies)
```



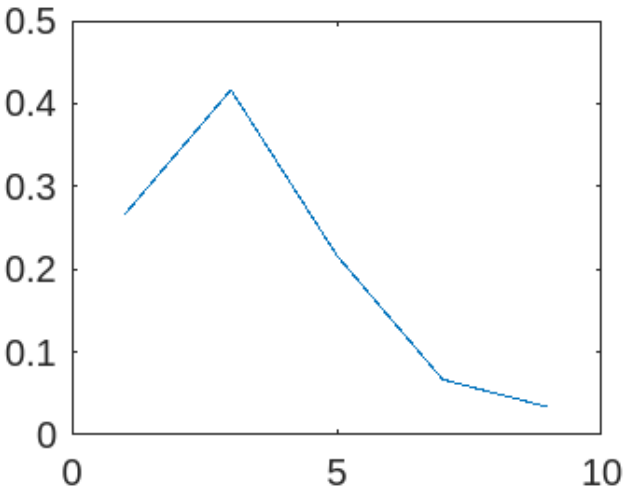
Relative frequency is found by dividing each classes frequency by total frequency. First total frequency is calculated, then relative frequencies are. The histogram is plotted at the end.

```
% relative frequency histrogram

totalFreq = 0;
for i=1:length(frequencies)
    totalFreq = totalFreq + frequencies(i);
end
relativeFreqs = frequencies/totalFreq;

figure(2);
bar(classMarks, relativeFreqs);
```



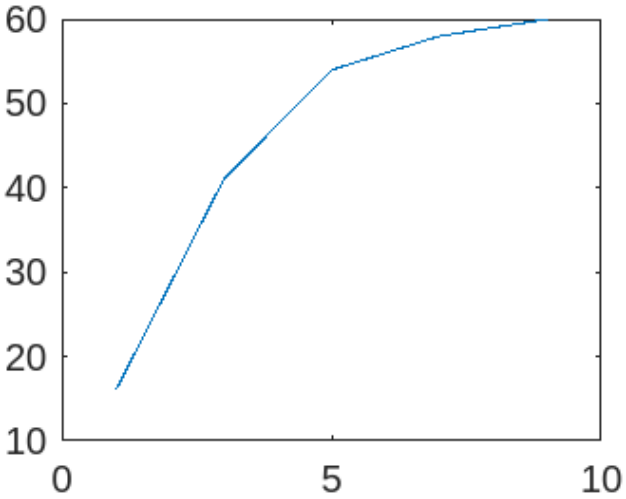
A relative frequency curve is plotted:

```
% relative frequency distribution

figure(3);
plot(classMarks, relativeFreqs);
```



The cumulative frequency curve is plotted:

```
% cumulative frequency distribution

cumulativeFreqs = zeros(length(frequencies), 1);
for i=1:length(frequencies)
    if i == 1
        cumulativeFreqs(i) = frequencies(i);
    else
        cumulativeFreqs(i) = cumulativeFreqs(i-1) + frequencies(i);
    end
end

figure(4);
plot(classMarks, cumulativeFreqs);
```



And here are the numerical values of frequencies:

frequencies	relativeFreqs	cumulativeFreqs
16	0.26667	16
25	0.41667	41
13	0.21667	54
4	0.066667	58
2	0.033333	60

2 Question 1.b

Here are mean, median, mode, standard deviation, variance, and skewness coefficient is calculated by MATLAB. **The mean is found as 3.366667:**

```
%% B

% mean
mean = sum(frequencies.*classMarks) / totalFreq;
fprintf("mean is %f \n", mean)
```

To calculate median, those steps are taken:

1. Median location is detected.
2. The index of the class, which contains median, is found.
3. Using that class index, other parameters are calculated.
4. The median is calculated.

The median is found as 3.120000:

```
% median

% The median is located at the middle.
medianPoint = totalFreq/2;
if rem(medianPoint, 2) == 0
    % if location index is even its index is between two values
    medianPoint = medianPoint + 0.5;
end

tempFreqSum = 0;
classContainingMedian = -1;
% The index of the class which contains median is found
for i=1:length(frequencies)
    tempFreqSum = tempFreqSum + frequencies(i);
    if medianPoint < tempFreqSum
        classContainingMedian = i;
        break
    end
end

L = classBoundaries(classContainingMedian); % Lower class boundary of the class containing
the median
N = totalFreq; % the number of items in data
freqSum = 0;
for i=1:length(frequencies)
    if (i >= classContainingMedian)
        break
    end
    freqSum = freqSum + frequencies(i);
end
freqMedian = frequencies(classContainingMedian); % sum of frequencies of all classes lower
than median class
C = abs(classBoundaries(classContainingMedian, 1) - classBoundaries(classContainingMedian,
2)); % size of median class interval

median = L + ((N/2)-freqSum)*(C/freqMedian);
fprintf("Median is %f \n", median);
```

The mode is found as 2.857143:

```
% mode

[greatestFreq, index] = max(frequencies);
delta1 = abs(greatestFreq - frequencies(index-1));
delta2 = abs(frequencies(index+1) - greatestFreq);
mode = L + (delta1 / (delta1+delta2)) * C;
fprintf("Mode is %f \n", mode);
```

The standard deviation is found as 2.016322:

```
% standart deviation

fx2 = sum(frequencies.*(classMarks.*classMarks));
standartDeviation=sqrt(fx2/totalFreq-mean*mean);
fprintf("Standart Deviation is %f \n", standartDeviation);
```

The variance is found as 4.065556:

```
% variance

variance = standartDeviation*standartDeviation;
fprintf("Variance is %f \n", variance);
```

The skewness coefficient is found as 0.442793:

```
% skewness

fx3 = (1/totalFreq)*sum((classMarks - mean).^3);
skewness = fx3 / standartDeviation^3;
fprintf("Skewness Coefficient is is %f \n", skewness);
```

3 Question 1.c

In this section, calculations are made by hand to compare with values found in MATLAB. Whereas mean, standard deviation and variance, calculated by hand, are almost as same as values, calculated by MATLAB, the skewness coefficient is found slightly different. Difference between two calculations are

$$0,442_{matlab} - 0,438_{hand} = 0,004$$

mean value

$$\frac{16.1 + 25.3 + 13.5 + 4.7 + 2.9}{16 + 25 + 13 + 4 + 2} = 3,36$$

std

$$S = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{16(1-3,36)^2 + 25(3-3,36)^2 + 13(5-3,36)^2 + 4(7-3,36) + 2(9-3,36)^2}{59}}$$

$S = 2,016$

$S^2 = 4,06$ (variance)

Skewness Coefficient

$$C_S = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{x_i - \bar{x}}{S^3} \right)^3$$

$$= \frac{1}{60} \frac{(16(1-3,36)^3 + 25(3-3,36)^3 + 13(5-3,36)^3 + 4(7-3,36)^3 + 2(9-3,36)^3)}{S^3}$$

$= 0,0438$