Chapter 1
Data representation

Statistics is defined as a collection of methods for collecting, organizing and representing data There two types of statistics:

1- Descriptive statistics : which describe data or represent it, data can be described using frequency count, average, modes, median and standard deviation
2- Inferential statistics: trying to guess the population parameters using the given sample . it performs hypothesis testing and mares prediction .

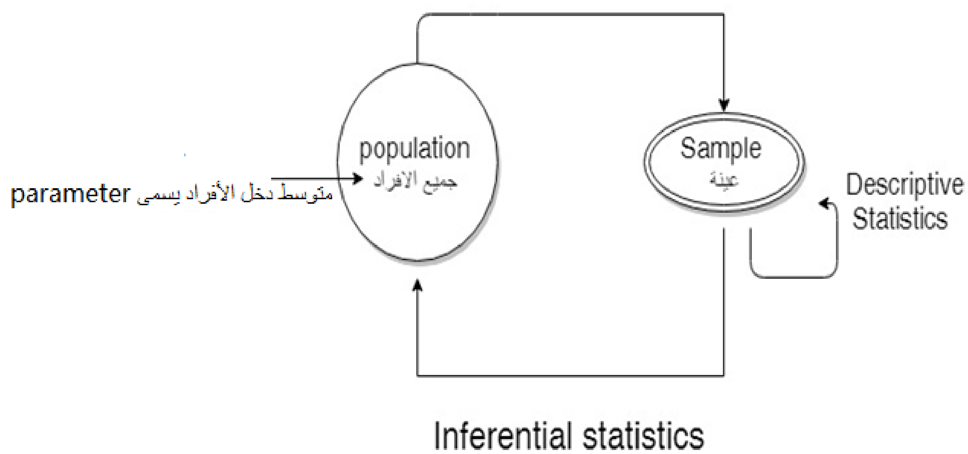


Figure 1.1: Scope of Statistics

Figure 1.1 shows the scope of statistics, which includes population, sample, parameters and statistic.

The population : the collection all elements (scores, persons, measurements, …. Etc.) to be studied

Sample : subset of the population

Parameter : characteristic or measure obtained from the population .

Statistic : is a characteristic or measure obtained from the sample .

In the field of statistics we have a collection or raw data, the data is collected from measurements, experiment results. As shown in Figure 1.2, The experiment study started by collecting data either from sensors (weather, salaries) or a result from another experiment. The data collected at this stage is called raw data. If the data is huge to commit study we have to take a sample of the data.

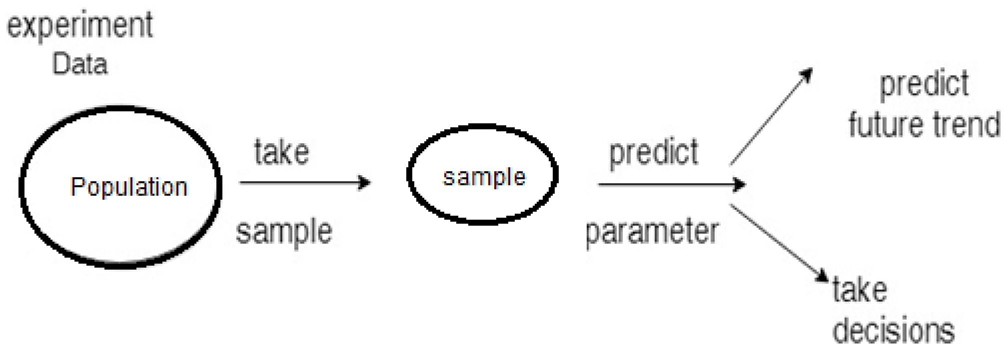


Figure 1.2: stages of making statistical experiments

1.1 Types of Data

The data can be qualitative such as “female”, “male”, or “Egyptian”, “Saudi” etc. generally qualitative data can’t be represented with numbers. Quantitative data is the data that can be represented numerically such as 0, 5, 21, 300, 0.12, etc. Quantitative data can be discrete or continuous (Bluman 2004)

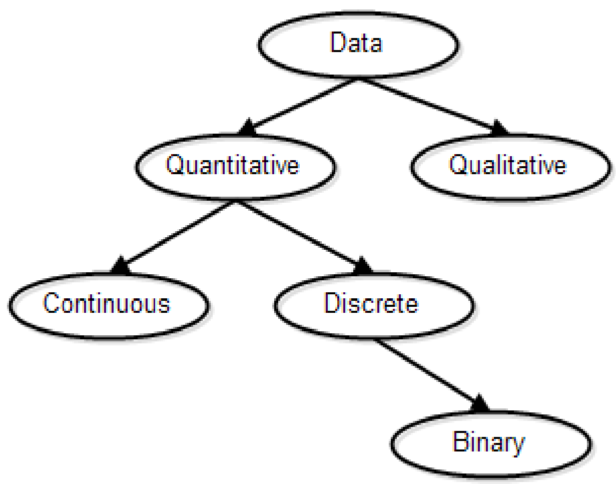


Figure 1.3: Types of data

Discrete Data : The data that its value is undividable such as

- Number of children in a family
- Number of students
- Number of crimes
- Number of cars sold in a day

Generally , we wouldn't expect to find 2.5 children or $\frac{1}{2}$ of a car sold in aday.

Binary date is the data that can have one of two values, such as of 1 or 0, Binary data is discrete data type.

Continuous Data **:** the observation that can take any value within finite or infinite interval real numbers, in other words it contains fractions.

Examples are :

- Weight
- Height
- Time to run 500 meters
- Age

1.2 Organization of Data

The statistician must organize data into meaningful way and present it so readers Can understand the distribution of data .

1.2.1 – Frequency table

Counting the number of repetition of specific value in the data.

Example : draw frequency table for the.

1 2 6 7 2 6 5 7 1 5 5 6 5 5 2

Solution :

Date	1	6	5	2	7
frequency	2	3	5	3	2

1.2.2 – Grouped Frequency table

Counting the frequency of repeating a value in specific range

Example : Draw the grouped frequency table of the following data with 5 classes .

12,22,18,9,25,31,28,19,22,27,32,14

Solution :

- Find the min and max values min 9, max 32 .
- wide the range over the class number
- $\frac{32-9}{5} = \frac{range}{\#\, classes} = \frac{23}{5} = 4.6$
- get the ceiling of the value width of class = $ciel\left(\frac{range}{\#\, classes}\right)$
- set the classes aB show in in the table each step increase 5

Class	**Frequency**
9 – 13	2
14 – 18	2
19 – 23	3
24 – 28	3

29 – 33	2

1.2.3 – Histogram

Displaying the frequency table with touching and non-overlapping bars .

Example : Use histogram to represent the above grouped frequency table .

Solution :

# Class	Frequency
9 – 13	2
14 – 18	2
19 – 23	3
24 – 28	3
29 – 33	2

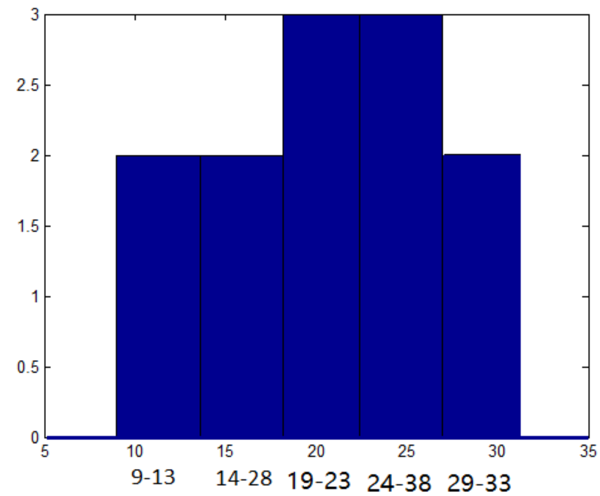


Figure 1.4 Histogram

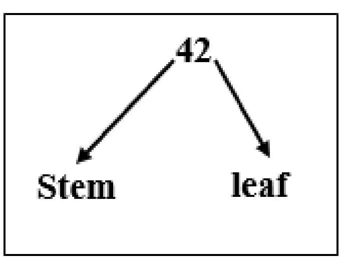
Frequency histogram with MATLAB use the following commands

```
>> a=[12,22,18,9,25,31,28,19,22,27,32,14]
>> hist ( a,5 )
```

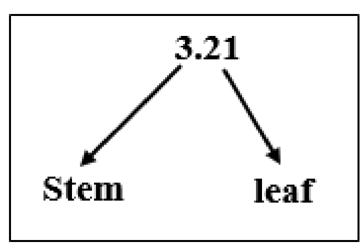
1.2.4 – Stem and leaf plot

Divide the data into two parts, STEM OF leaf list the Stems in Column and leaf's in rows .

Discrete



Continuous



<u>Example</u> : Draw Stem of leaf for the following Date .

90 70 70 70 75 70 65 68 60

74 70 95 75 70 68 65 40 65

Answer :

4	0
5	
6	055588
7	000000455
8	
9	05

Notice : 9 05 means there are data value, 90 and 95 in the data .

Example: Let's consider a random sample of 20 concentrations of calcium carbonate (CaCO3) in milligrams per liter.

130.8	129.9	131.5	131.2	129.5	132.7	131.5	127.8	133.7
132.2	134.8	131.7	133.9	129.8	131.4	128.8	132.7	132.8
131.4	131.3							

Create a stem-and-leaf plot of the data.

Solution: The following figure shows the resulting stem and leaf plot for the above continuous data

127	8
128	8
129	589
130	8
131	2344557
132	2778
133	79
134	8

1.3 Central tendency

The data by its nature if often tend to be concentrated around the center of data, the measure of this type is called measure of location such as the following measures.

- Mean
- Median
- Mode

1.3.1 : Arithmetic mean

The mean of measurements is the sum of all measurements divided by number of measurement

$mean\ \bar{x} = \frac{\sum xi}{n}$

n : is the number of measurements

xi : is the ith measurement

Note: notice that μ is the population mean while $\bar{x}$ is the sample mean.

Example: find the mean of following sample data, 2, 9, 11, 5, 6

Solution: $\overline{x} = \frac{\sum xi}{n} = \frac{2+9+11+5+6}{5} = \frac{33}{5} = 6.6$.

The following MATLAB command can be used to get the arithmetic mean

```
>> mean(a)
```

1.3.2 : Median

The median of a set of measurements is the middle measurement after sorting the data

The position of median value is $[\frac{n+1}{2}]$

Example: find the median of 2,4,9,8,6,5,3

Solution:

1- Sort the numbers 2,3,4,5,6,8,9
2- Position of median $= \frac{7+1}{2} = 4$

So the median is the fourth number which is 5. The median is obtained directly if the number of elements in the data (n) is odd.

Example: Find the median of 2,4,9,8,6,5

This an even problem, $n = 6$ so the median position will be between the third and fourth positions.

1- Sort the numbers 2,4,5,6,8,9
2- Position $= \frac{n+1}{2} = \frac{6+1}{2} = 3.5$

$$2,4,5, position, 6 , 8 , 6$$

The median is $= \frac{5+6}{2} = \frac{11}{2} = 5.5$

1.3.3 mode:

Mode measures the data which occurs most frequently

Example : the set $2, 4, 9, 8, 8 , 5, 3$ has mode of 8

Example : the set $2, 2, 9, 8, 8, 5, 3$ has mode of 2, 8 (bimodal) .

Example : the set $2, 9, 3, 6$ has mode of 2 the first number .

1.3.4 Removing Outliers (إزالة القيم الشاذة)

The median filter: the median filter collects 3 by 3 samples of the data and finds their median.

Example: filter the following data with median filter $2 , 4 , 10 , 5 , 3$, notice that 10 is an outlier

Solution:

find the median of (2 4 10)

$$sort \Rightarrow 2, 4, 10 \Rightarrow median = \underline{4}$$

find the median of (4 , 10 , 5)

$$sort \Rightarrow 4, 5, 10 \Rightarrow median = \underline{5}$$

find the median of (10 , 5 , 3)

$$sort \Rightarrow 3\,,5\,,10 \Rightarrow median = \underline{5}$$

The resulting signal is $2, 4, 5, 5, 3$, notice that the outlier value (10) is filtered out.

Note that median filter removes sharp notches (outliers)

Trimming: it is obtained by removing the highest and the lowest part of the data. For this purpose the data must be sorted first. For example 10% trimmed data is obtained by removing 10% of the data from both highest and lowest values.

Example: The given data is measurements taken as an output from experiment.

3.4, 2.5, 4.8, 2.9, 3.6, 2.8, 3.3. 5.6, 3.7, 2.8, 4.4, 4.0, 5.2, 3.0, 4.8

Assuming the measurements are a random sample, compute the 20% trimmed data for the data set:

Solution: sample size is n=15, 20% of 15= 0.20 * 15=3, So we remove the first and last 3 values after sorting them,

The sorted data is : 2.5, 2.8, 2.8, 2.9, 3.0, 3.3, 3.4, 3.6, 3.7, 4.0, 4.4, 4.8, 4.8, 5.2, 5.6

The trimmed data is: 2.9, 3.0, 3.3, 3.4, 3.6, 3.7, 4.0, 4.4, 4.8

MATLAB (A)

A.1: Basic operations .

Basic arithmetic and logic operations can be open done with MATLAB Such as

$\gg \mathbf{16/10}$ $\gg \mathbf{30} * \mathbf{5}$ $\gg \mathbf{30 - 300}$ $\gg \mathbf{500 + 7}$ $\gg \mathbf{30} * (\mathbf{500 + 7})$

In addition, a variable Can be defined. Then used such as

$\gg \boldsymbol{a} = \mathbf{10};$

$\gg \boldsymbol{z} = \boldsymbol{a} * \mathbf{30}$

$\gg \boldsymbol{q} = \boldsymbol{z} * (\boldsymbol{a} + \mathbf{5})$

You can ignore the result (not display it) by adding ";" to the end of the command such as

$\gg \mathbf{16/10};$

A.2:Vector representation :

Vector can hold 1D data such as voice signal or data transferred on a network, etc.

Vector can be represented as

$\gg \boldsymbol{a} = [\,\mathbf{5\ \ 10\ \ 15\ \ 30\ \ 55\ \ 16\ \ 27\ \ 300\ \ 5}\,];$

Or

$\gg \boldsymbol{a} = [\,\mathbf{5\ \ 10\ \ 15\ \ 30\ \ 55\ \ 16\ \ 27\ \ 300\ \ 5}\,];$

Regular vector Can be created as

$\gg \boldsymbol{s} = [\mathbf{1:10}]$

The result must be **1 2 3 4 5 6 7 8 9 10**
The vector is increased by 1, you can change .
The default incremental step to 2 such that
$\gg s = [\mathbf{1} : \mathbf{2} : \mathbf{10}]$
The result should be **1 3 5 7 9**
Random vector can be created with the command **rand.** it creates random values between 0 and 1 this default can be changed to create random values between 0 and 100 as follow
$\gg \mathbf{ceil}\,(\,\mathbf{rand}\,(\,\mathbf{1}, \mathbf{10}\,) * \mathbf{100}\,)$
Notice that **rand** command creates single random number
$\gg$ **rand**
While a vector of random numbers between 0 and 1 and the length of the vector is 10 Can be created with
$\gg \mathbf{rand}\,(\,\mathbf{1}\,, \mathbf{10}\,)$
The floating numbers can be removed with the function **ceil** for example *ceil (0.11) = 1* and *ceil (9.2) = 10*
the vector can be plotted and represent as follow:
$\gg \mathbf{a} = \mathbf{ceil}\,(\mathbf{rand}\,(\,\mathbf{1}, \mathbf{10}\,) * \mathbf{100}\,)$
$\gg \mathbf{plot}\,(\mathbf{a})$
$or \qquad \gg \mathbf{figure}, \mathbf{plot}\,(\mathbf{a})$

The following command also creates random numbers with uniform distribution, notice we used **fix** instead of **ceil.**
$\gg \mathbf{a} = \mathbf{fix}(\mathbf{rand}(\mathbf{1}, \mathbf{10}) * \mathbf{100})$
>> hist(a,6)
The following command can create random numbers with normal distribution.
>> a=fix(normrnd(70,10,1,20)
>>hist(a,6)
Where 70 is the center of data or the mean, 10 is the fluctuation range or the standard deviation, and 20 is the number of data points. Notice that the second figure displays the bell shaped distribution or histogram of heights frequencies.

A.3:Mean

The mean is the center of data which data values fluctuates around it, mean can be easily found with the **mean** MATLAB command

$\gg$ **a = fix (rand (1,10) * 100)**

$\gg$ **m = mean (a)**

For a signal a , it can be filtered using the following MATLAB command.

$\gg$ **sig 1 = conv(a, [0.333 0.333 0.333],' same')**

$\gg$ **plot (1 : 10, a , 1 : 10 , sig1)**

A.4:median

The median is the number located at the center of data, in the same manner as mean, median can be obtained as :

$\gg$ **a = fix(rand (1 ,10) * 100)**

$\gg$ **med = median (a)**

A signal can be filtered with median filter using the following Command.

$\gg$ **sig2 = medfilt1(a, 3)**

$\gg$ **plot (1 : 10 , a , 1: 10, sig2)**

A.5 mode:

Mode can be obtained with the **mode** MATLAB Command.

$\gg$ **a = fix (rand (1 ,10) * 100)**

$\gg$ **mode (a)**

1.4: Measures of Dispersion (مقاييس التشتت في البيانات)

Mean and median measure tells useful information about central tendency, but they lack the ability to describe data variance. for example, data may have the same mean but totally different variance.

Example 1 :

Calculate mean for the following two sets of data .

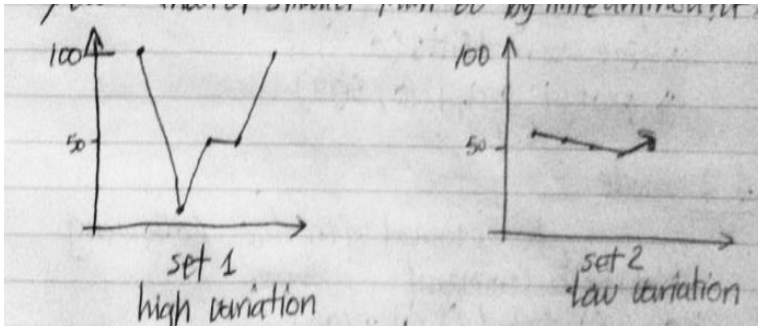
Set 1 : 100 10 50 50 90

Set 2: 62 61 54 58 60

Mean of set 1 = $\frac{100+10+50+90}{5} = \frac{300}{5} = 60$

Mean of set 2 = $\frac{62+61+59+58+60}{5} = \frac{300}{5} = 60$

Set 1 & set 2 have the same mean which is 60. But they have different variance . set 1 has very large values such as 100 & 90 . and very small values such as 10 . set2 has little variance since the values is greater than or smaller than 60 by little amount .



Set 1 gets far away from the central or mean value, but set 2 fluctuate around the center value.
There are many methods that can be used to measure the variations, each method has its own advantages and disadvantages.
One of these methods is the range.

1. 4.1: Range

range is simply the difference between the largest and smallest values in the dataset .

$$\textbf{Range}\ (\boldsymbol{x1}, \boldsymbol{x2}, \boldsymbol{x3}, \boldsymbol{x4}) = \textbf{max}(\boldsymbol{x1}, \boldsymbol{x2}, \boldsymbol{x3}, \boldsymbol{x4}) - \textbf{min}(\boldsymbol{x1}, \boldsymbol{x2}, \boldsymbol{x3}, \boldsymbol{x4})$$

<u>Example</u> : find the ranges of the data in example 1

Set 1 : 100 10 50 50 90

Set 2: 62 61 59 58 60

Solation :

$$\min(100\ ,10\ ,50\ ,50\ ,90 = 10$$
$$\max(\ 100\ ,10\ ,50\ ,50\ ,90\ ) = 100$$

Set1 range = 100 – 10 = 90

$$\min(62\ ,61\ ,59\ ,58\ ,60 = 58$$
$$\max(\ 62\ ,61\ ,59\ ,58\ ,60\ ) = 62$$

Set2 range = 62 – 58 = 4

Set1 has high rang range which represents high variance and set 2 had low range (4) which represents low fluctuations.

Hint Range is a good measure for variability, but it is very weak if the data had outliers (قيم شاذة) the outliers affects the range but range in this case , will not express real variability.

القيم الشاذة تؤثر في المدى. على سبيل المثال إذا اردنا معرفة مدى الرواتب في مصنع من المصانع، وكان حوالي 100 عامل في المصنع يتقاضون رواتب بين 3 آلاف ريال وأربعة آلاف ريال. بذلك سيكون المدى بسيط وهو ألف ريال وهو مدى منطقي. أما إذا وجد مدير للمصنع يتقاضى 50 الف ريال. وهو الوحيد في المصنع الذي يتقاضى هذا الراتب. فإذا حسبنا المدى سيتأثر بالقيمة الشاذة وهي قيمة الراتب 50 الف. وسيصبح 47 ألف ريال. وهو غير منطقي حيث أن التذبذب الخاص بالبيانات ليس مداه 50 ألف.

1 . 4 . 2 : Variance and Standard Deviation

if (x1 , x2 , x3) is the data, then the variance an be calculated by

$$S^2 = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + (x_3 - \overline{x})^2}{n-1}$$

Where $\overline{x} = \frac{x_1 + x_2 + x_3}{2}$ which is the data average or mean. and n is the number of data values which is 3 in this case .

The sample variance is S^2

The population variance is σ^2

The variance is simply the average value of the distances between each value in the dataset and data-set central value. or the average value or the fluctuations variance is expressed mathematically as:

$$S^2 = \frac{1}{n-1} \sum_{1=1}^{n} (x_i - \overline{x})^2$$

The variance is fine measure for describing variability, but the resulting value is squared, for example if the data set is heights of persons, the resulting values are not reflecting heights, it represent the square of the height.

<u>Example:</u> find the variance marks of the following sets which represents the degrees of STAT class students.

Set 1 : 100 10 50 50 90

Set 2 : 62 61 59 58 60

Set1 variance ($\overline{x} = 60$)

$$s_1^2 = \frac{1}{5-1} [(100-60)^2 + (10-60)^2 + (50-60)^2 + (50-60)^2 + (90-60)^2]$$

$$s_1^2 = 1300$$

Set2 variance ($\overline{x} = 60$)

$$s_1^2 = \frac{1}{5-1} [(62-60)^2 + (61-60)^2 + (59-60)^2 + (58-60)^2 + (60-60)^2]$$

$$s_2^1 = \frac{4+1+1+4+0}{4} = \frac{10}{4} = 2.5$$

So, set1 has higher variances (1300) them set2 (2.5)

<u>Drawback</u>: The problem is that 1300 is not a student mark, it doesn't reflect the same range of the data set , no one can get 1300 in the STAT Class !!

To solve the above problem, the standard deviation is used, the standard divination is simply the square root of the variance .

$$s = \sqrt{s^2}$$

For set 1 $s = \sqrt{1300} = 36.056$

For set 2 $s = \sqrt{2.5} = 1.5811$

Now we can easily say that the deviation of date from the center is 36 marks .

Another formula of variance

$$S^2 = \frac{\sum_{1=1}^{n} x_i^{\ 2} - n\bar{x}^2}{n-1}$$

So set1 variance $(\overline{x} = 60)$

$$s^2 = \frac{((100)^2 + (10)^2 + (50)^2 + (50)^2 + (90)^2 - (5 \times (60)^2)}{5-1}$$

$$= \frac{10000 + 100 + 2500 + 2500 + 8100 - (5 \times 3600)}{4}$$

$$= \frac{5200}{4} = 1300$$

And set2 variance

$$s^2 = \frac{((62)^2 + (61)^2 + (59)^2 + (58)^2 + (60)^2 - (5 \times (60)^2)}{5-41}$$

$$= \frac{10}{4} = 2.5$$

Exercise: Compute the variance and standard deviation of the following datasets

Set 1 10 21 33 53 54

Set 2 34 36 35 33 34

Explain the difference between the results, justify your answer.

1. 4. 3: Box Plot

To draw box plot, the inter – quartile range must be extracted from the data. The range of the middle of 50% score is the interquartile range. or

$$IRQ = Q3 - Q1$$

Where Q3 is the median of the right half of the data. and Q1 is the median of the left half of the data.

Example 1: Given the following dataset calculate the interquartile-range.

$$2, 3, 5, 6, 7, 8, 9, 10$$

Ans : the median = $\frac{6+7}{2} = 6.5$

Q1 is the median of 2 , 3 , 5 , 6 which is $\frac{3+5}{2} = 4$

Q3 is the median of 7 , 9 , 9 , 10 which is $\frac{9+9}{2} = 2$

Example 2: Calculate the interquartile range of $1, 3, 5, 6, 7, 8, 8$

Ans **:** median = 6

Q1 is the median of 1 , 3 , 5 which is 3

Q3 is the median of 7 , 8 , 8 which is 8

Drawing the box plot: the median, Q1 and Q3 can be used for drawing the box plot

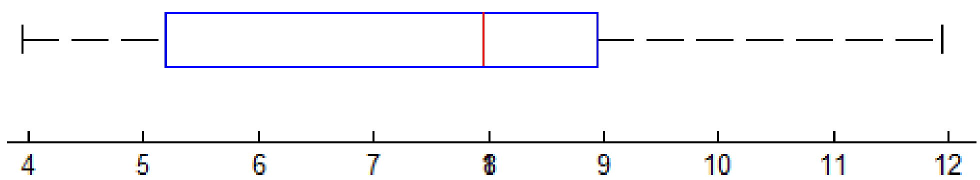
Example: Draw the box plot for the following data

$$4, 4, 5, 6, 8, 8, 9, 9, 9, 10, 12$$

Median = 8

$$Q1 = 5.5 \quad and \quad Q3 = 9$$

The box plot is shown in the figure .



MATLAB (B)

Try the following - exercises on MATLAB and Octave

$\gg$ **a** = [**100 10 50 50 90**]

$\gg$ **b** = [**62 61 59 58 60**]

$\gg$ **range** (**a**)

$\gg$ **range** (**b**)

$\gg$ **var** (**a**)

$\gg$ **var** (**b**)

$\gg$ **std** (**a**)

$\gg$ **std** (**b**)

>> boxplot(a)

range is function for getting range of data.

var is function to get variance of data.

std is the standard deviation.

boxplot is function for drawing the boxplot figure.

Exercises1

Question 1: for the following sample points 9,2,7,11,14,7,2,7,2,7,9

- Draw the frequency table
- Draw the histogram, number of classes = 5
- Calculate sample mean
- Calculate sample median
- Get sample mode
- Calculate sample Range

Question 2: A sample of 20 measurements is shown here:

2.6	3.4	2.1	3.2	4.2	3.6	2.8	3.8	1.7	3.9
2.2	1.2	5.6	3.9	2.5	4.1	3.0	2.3	2.7	1.9

- Use the data in the table to make a stem-and-leaf plot.
- Drive the frequency table from the data assuming number of classes = 5
- Draw the Histogram for the data.

Question 3: The ages of 5 randomly selected members of a club are as follows: 42, 52, 57, 63, 51

- The sample mean is (……….)
- The sample median is (……..)
- The sample Range is (…….)
- The sample Standard Deviation is (……)

Question 4: The following table represents the degrees out of 100 for Statistics and Math exams

STAT 301	74	68	62	69	67
MATH 333	58	98	38	78	68

(a) Calculate the sample Mean for the marks of each subject;
Can you differentiate between the two sets using Mean? Why?

(b) Calculate the sample Standard Deviation for the marks of each subject,
Can you differentiate between the two sets using Standard Deviation? Why? in what situation will the standard deviation be zero?

Question 5: Connect the following statements with the correct expression.

(a)	Eye color : blue, brown, hazel, green, etc.	(……)	Discrete
(b)	Number of children, Number of students,	(……)	Binary Data
(c)	Very unhappy, unhappy, neutral, happy,	(……)	Nominal Data
(d)	Gender: Male, Female	(……)	Continuous
(e)	Height, Temperature, Age	(……)	Ordinal Data

Question6: The following are the blood group of sample patients who attend clinic A.

A,	B,	O,	AB,	B,	A,	O,	O,	AB,	B
B,	B,	A,	O,	O,	AB,	B,	O,	B,	A
AB,	A,	O,	A,	A,	B,	O,	A,	A,	B

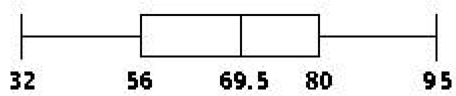
a. Construct the frequency table of the above data
b. Draw the histogram of the above data.

Question7. Draw a box plot for the following data set:

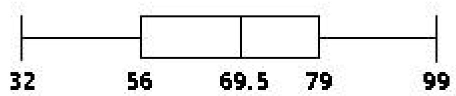
4.3, 5.1, 3.9, 4.5, 4.4, 4.9, 5.0, 4.7, 4.1, 4.6, 4.4, 4.3, 4.8, 4.4, 4.2, 4.5, 4.4

Multiple Choices Questions: Choose the correct answer for the following questions.

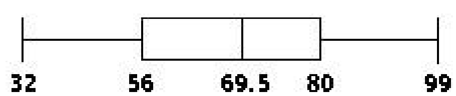
1) The number of students entering college in a certain year is 621.
a) Continuous b) Discrete
2) Calculate the Standard Deviation for the following data which was sampled from a large population: 4,10,12,2,15,5
a) 5.099 b) 26 c) 0 d) 15.12
3) The ages (in years) of the eight passengers on a bus are listed below.
10 7 26 16 21 43 40 30 Find the median age.
a)23.5 yr b)21 yr c)26 yr d)24.5 yr
4)The distances (in miles) in the past week by each of a company's sales representatives are listed below. 107 114 214 230 436 445 Find the mean.
a)214 miles b)220.50 miles c)257.67 miles d)230 miles
5)Find the mode for the given sample data. -20 -43 -46 -43 -49 -43 -49
a)-49 b)-46 c)-41.9 d)-43
6)The test scores of 32 students are listed below. Construct a box plot for the data set.
32 37 41 44 46 48 53 55 57 57 59 63 65 66 68 69 70 71 74 74 75 77 78 79 81 82 83 86 89 92 95 99
A)



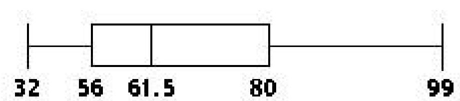
B)



C)



D)



7) Suppose S={1,2,3,4,5,6} is the space for an output of tossing a die, which of the following is a true event

a) E={-1,2,1} b) E={1,2,7} c) E={1,2,6} d) E={1,0,7}

8)The temperatures in 7 different cities on New Year's Day are listed below.

26 29 33 59 67 68 78

Find the median temperature.

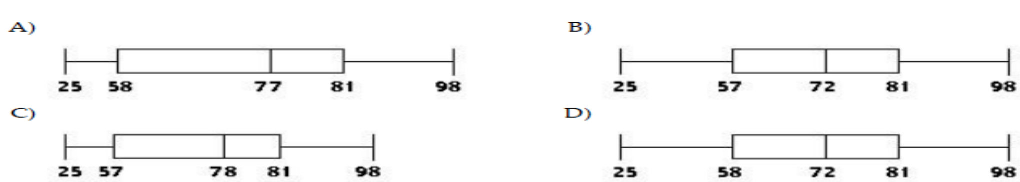
A)33°F B)59°F C)51°F D)67°F

9) Find the variance for the given data. 6.6 8.5 4.6 1.7 2.4

A)8.08 B)7.98 C)6.46 D)12.80

10) The test scores of 40 students are listed below. Construct a boxplot for the data set. D

25	35	43	44	47	48	54	55	56	57
59	62	63	65	66	68	69	69	71	72
72	73	74	76	77	77	78	79	80	81
81	82	83	85	89	92	93	94	97	98

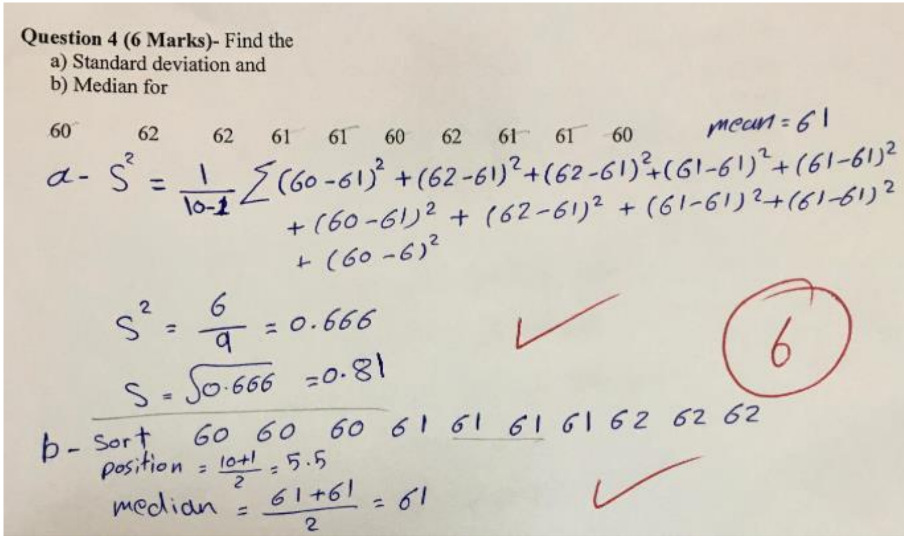


11) If the sample size is 9 and the standard deviation is 7 then the variance is:

a) 49 b) 2.6 c) 2 d) 81

Solution Key

7 (c) 8 (b) 9 (a) 10 (d) 11 (a)



Question 4 (6 Marks)- Find the

a) Standard deviation and

b) Median for

60 62 62 61 61 60 62 61 61 60 mean = 61

a- $S^2 = \frac{1}{10-1} \sum (60-61)^2 + (62-61)^2 + (62-61)^2 + (61-61)^2 + (61-61)^2 + (60-61)^2 + (62-61)^2 + (61-61)^2 + (61-61)^2 + (60-6)^2$

$S^2 = \frac{6}{9} = 0.666$

$S = \sqrt{0.666} = 0.81$

6

b- sort 60 60 60 61 61 61 61 62 62 62

position $= \frac{10+1}{2} = 5.5$

median $= \frac{61+61}{2} = 61$