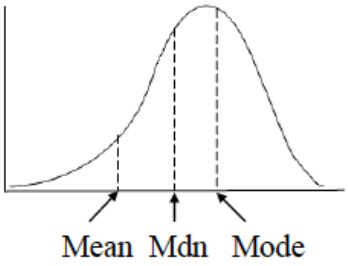
Chapter 2: Measures of Central Tendency

A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data.

There are three main measures of central tendency:

1) Mean
2) Mode
3) Median



1. Mean:

Mean is the arithmetic average of a set of data. Mean(denoted by $\bar{x}$ of a given set of data) can be found by using the following formula:

$$\bar{x} = \frac{Sum\ of\ the\ value\ of\ items}{Number\ of\ items}$$

Formula for calculating the Mean of a grouped frequency distribution:

$$\bar{x} = \frac{\sum fx}{\sum f}$$

Where,

f = frequency

x= variable

Example 1: Mean of an ungrouped frequency distribution

Find the mean of the following distribution:

x	f
2	34
3	24
4	53
5	42
6	23
7	25

Solution

The mode is of this distribution is 4 as it has the highest frequency

Example 2: Mode of a grouped frequency distribution

x	f
2-5	5
6-9	8
10-13	10
14-17	12
18-21	7
19-25	3

Solution

For this first we have to calculate the modal class, i.e; the class with the highest frequency.
Modal class = 8-11

$$Mode = L + \frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_3)} \times W$$

Where;
L= lower class boundary of the Modal class
f_m = frequency of the modal class
f_1 = frequency of the class preceding the modal class
f_3 = frequency of the class below the modal class
W = width of the modal class

Therefore, Mode of the above distribution will be:

$$\text{Mode} = 13.5 + \frac{(12-10)}{(12-10)+(12-7)} \times 4 = 15.1$$

3. Median

Median is the middle value in a set of data.

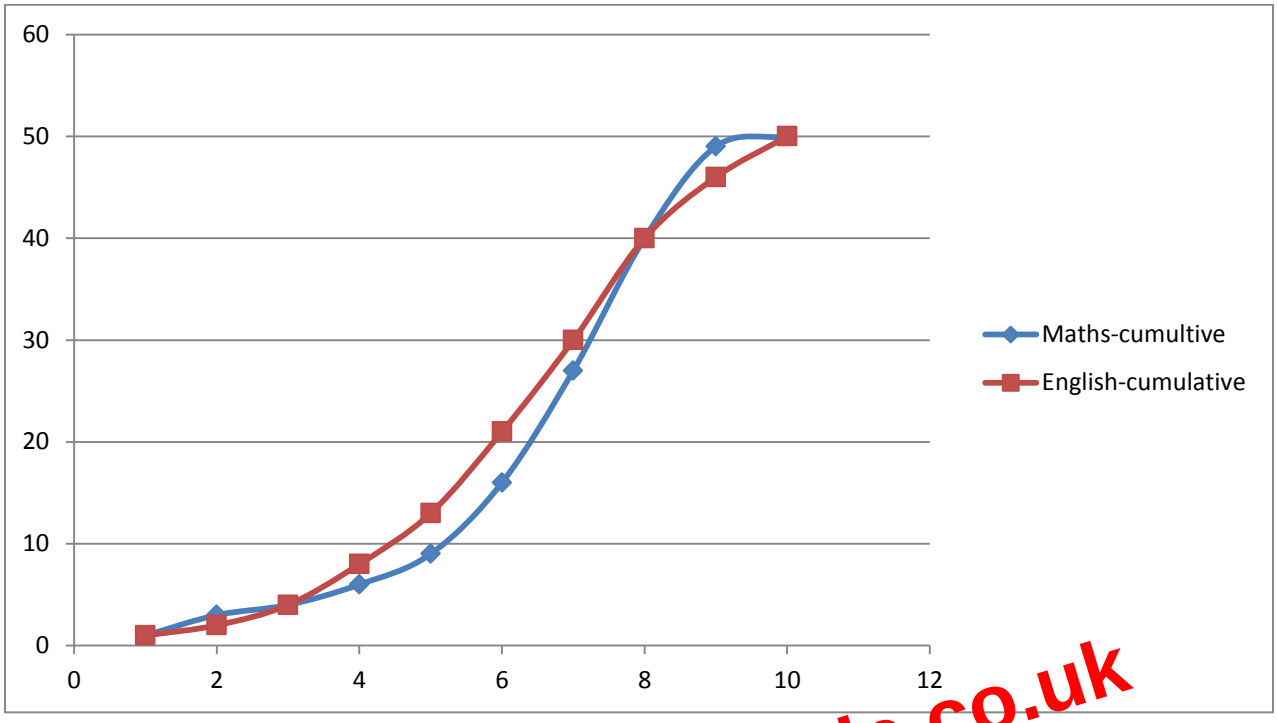
Example 1

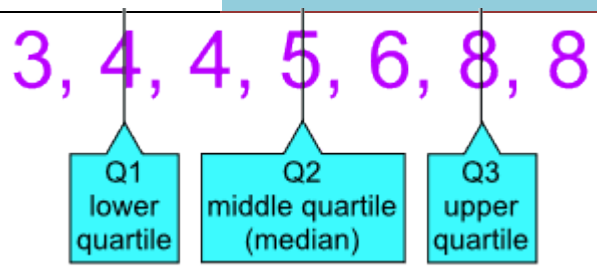
Calculate median of the following data:
4,2,6,3,7,2

Solution

First of all, arrange the data:

X = 2,2,3,4,6,7
Median = Sum of frequencies+1/2 th value = (6+1)/2 = 3.5th value,
= (3rd value + 4th value)/2 = 3.5





Chapter 4: Quartiles, Deciles and Percentiles

Quartiles, Deciles and Percentiles;

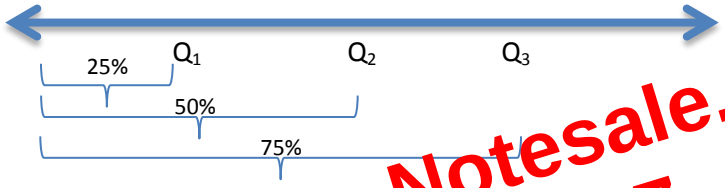
Quartiles

Quartiles are defined as three points that divide a set of data into four equal groups.

First Quartile (Q_1), also known as lower quartile, is the middle number between the smallest number and the median;

Second Quartile (Q_2) is the median; and

Third Quartile (Q_3), also known as upper quartile, is the middle number between the highest number and the median (Q_2).



Similarly, the values which divide the data into ten equal parts are known as Deciles, and are denoted by $D_1, D_2, \ldots D_9$, and the values which divide it into hundred equal parts are known as Percentiles, denoted y $P_1, P_2, \ldots P_{99}$.

Formulae	
First Quartile	$(\frac{n+1}{4})$ th observation
Second Quartile	$(\frac{n+1}{4}) \times 2$ th observation
Third Quartile	$(\frac{n+1}{4}) \times 3$ th observation
First Decile	$(\frac{n+1}{10})$ th observation
x Decile	$(\frac{n+1}{10}) \times x$ th observation
First Percentile	$(\frac{n+1}{100})$ th observation
x percentile	$(\frac{n+1}{100}) \times x$ th observation

Example

Find Q_1, Q_2,and Q_3 of the following data:
2,4,3,5,6,7,9,3,5,8,1

Solution

First of all, we have to arrange the data in ascending order:

Chapter 6: Standard Deviation and Variance

1.1. Variation/Dispersion

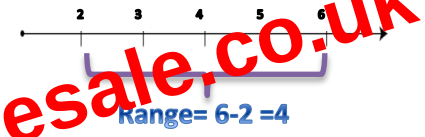
Variance measures how far a set of numbers is set out. There are a number of measures which can be used for providing this information, including:

- Range
- Mean Deviation
- Standard deviation
- Variance

1.2. The Range

The range is the difference between highest and lowest values of a set of data.

Range = Highest value – Lowest Value



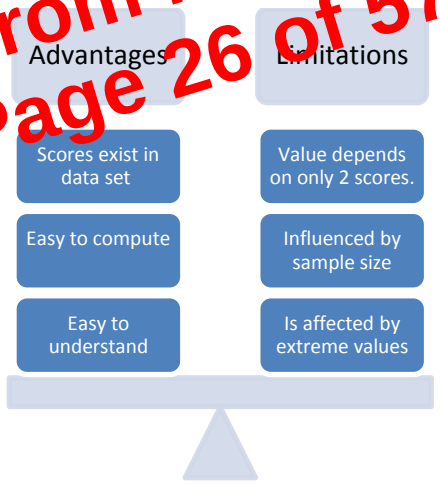


Figure 5 Advantages and Limitations of using range as measure of dispersion

Example

Calculate the mean and standard deviation of the following data:

X =3,4,6,4,5,3,5

Solution

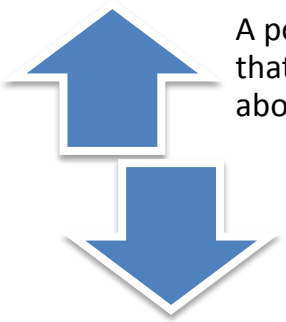
x	f	fx	fx2
3	4	12	36
4	3	12	48
6	5	30	180
4	3	12	48
5	2	10	50
3	2	6	18
5	1	5	25
30	**20**	**87**	**405**

Mean = $\frac{\sum fx}{\sum f}$ = 87/20 =4.35

Standard deviation = $\sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2} = \sqrt{\frac{405}{20} - 4.35^2}$ = 1.13

Standard Score

The z score is a measure of the number of standard deviations that an observation is above or below the mean.



A positive z-score indicates that the observation is above the mean

A negative z-score indicates that the observation is below the mean

For a set of data $x_1, x_2 ... x_n$,

Z score = $\frac{x-\bar{x}}{\sigma}$

Where,

$\bar{x}$ = mean of the data; and

b)

Score on 1st dice	Score on 2nd dice	Total score
1	5	6
2	4	6
2	5	7
2	6	8
3	3	6
3	4	7
3	5	8
3	6	9
4	2	6
4	3	7
4	4	8
4	5	9
5	1	6
5	2	7
5	3	8
5	4	9
6	1	
6	2	8
6	3	9

P(Greater than 5 but less than 10) = 19/36

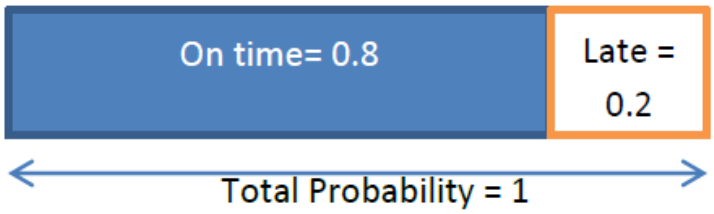
Activity

Continuing from the above example, find the probability of getting score:
a) Equal to 5
b) Less than or equal to 4
c) Greater than 10
d) Less than 3
e) Equal to 7

Instead of using a table to ascertain the total possible outcomes, tree diagram is used usually. It shows all possible outcomes and allows calculating probability. Each branch of the tree represents an outcome.

'Independent events' are those events in which the happening or non-happening of one event has not effect on other. If the outcome of one event affects the other, then these events are called **'Mutually Exclusive'.**

If you toss a coin, you can either get 'Heads' or 'Tails', but not both at the same time. Therefore, these are 'Mutually Exclusive Events'. If you toss a coin twice, then the outcome of 1st throw will have no effect on the 2nd one. So, these are 'Independent Events'. its probability will be independent of the 1st throw. Read the Example 1 again. The probability that 1 or 6 will come on the first throw is an example of 'Mutually Exclusive events', whereas the probability that 1 will come on the 1st throw and 6 on the 2nd

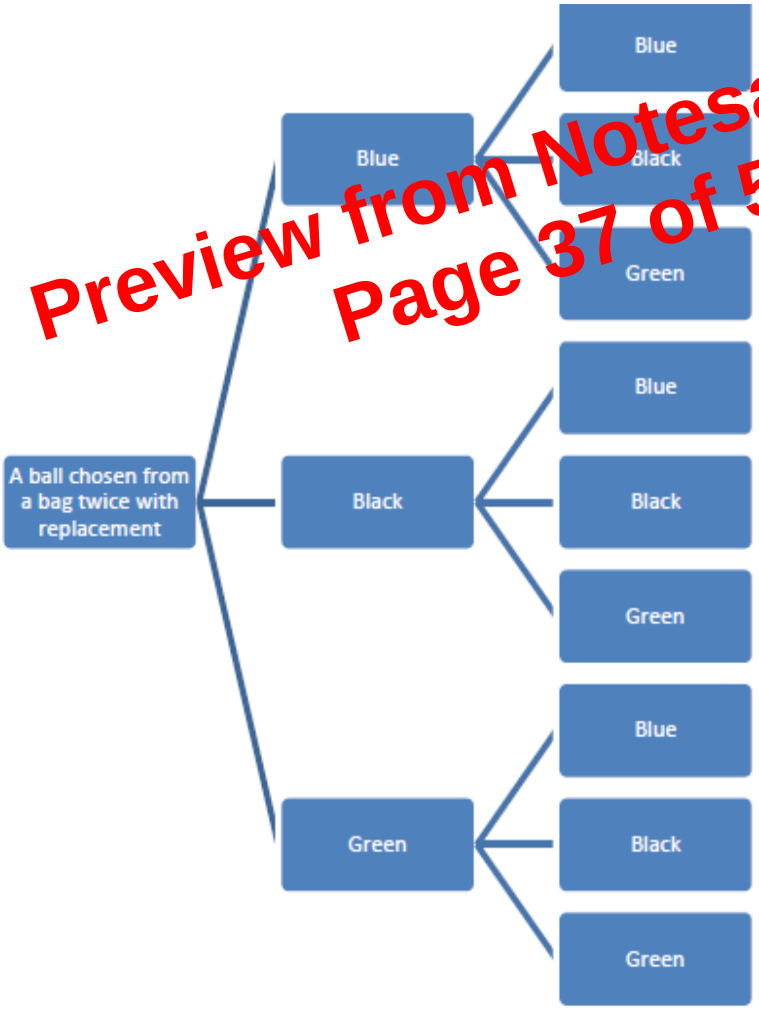


Example

A ball is chosen twice from a bag with replacement containing 6 Blue balls, 4 Black balls and 2 Green balls. Calculate the following probabilities.

a) P(Blue, Black)
b) Probability of getting at least 1 green ball
c) Probability of getting balls of same color both times

Solution



Centering Moving Average

In case of even moving averages, moving averages are centered, i.e adjusting them so that they may coincide with relevant year.

Example

Calculate moving average of the following data

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008
Sales	200	300	250	400	220	430	300	510	440

Solution

Year	Sales	4 year moving averages	Moving averages centered
2000	200		
2001	300		
2002	250	287.5	290
2003	400	292.5	308.75
2004	220	325	331.25
2005	430	337.5	351.25
2006	300	365	392.5
2007	510	420	
2008	440