

chapter

2

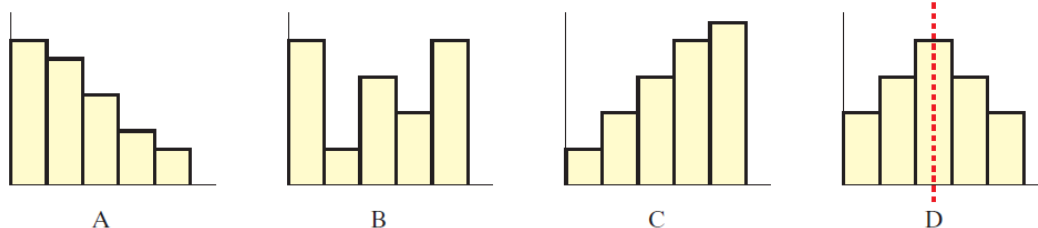
Statistics 1

Section 2.8 The shape of a distribution

PROJECT MATHS
Text & Tests 5
 LEAVING CERTIFICATE
 HIGHER LEVEL
 STRAND 1
 PROBABILITY & STATISTICS

90

In the previous section, we encountered histograms of various shapes. The diagrams below show four histograms, all with different shapes.



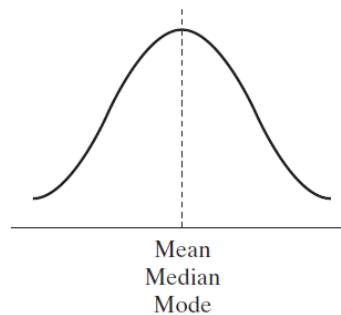
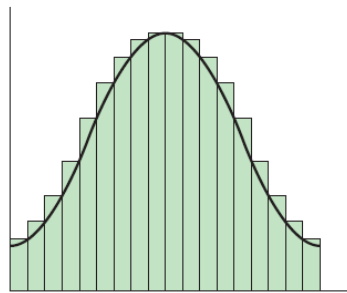
Only histogram D appears balanced or symmetrical as it has an axis of symmetry. The other three histograms are less balanced or **skewed** in some way.

Histograms are very useful when you want to see where the data lies and so get a clear picture of the shape of the distribution. For example, in histogram A above, we can see that most of the data is concentrated at the lower values. In histogram C, the data is concentrated at the higher values.

There are some shapes that occur frequently in distributions and you should be able to recognize and name them. The most common and frequently occurring shapes follow.

1. Symmetrical distributions

- › This distribution has an axis of symmetry down the middle. It is called a **symmetrical distribution**.



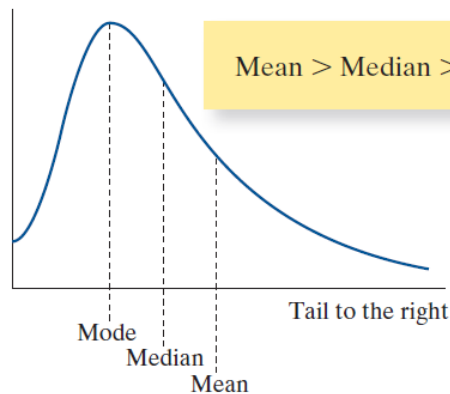
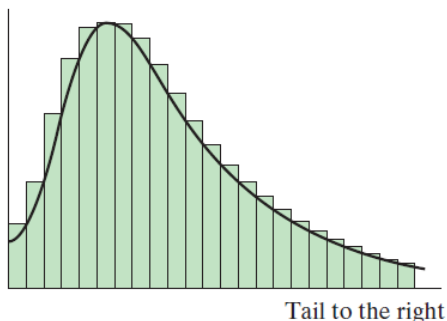
- › It is one of the most common and most important distributions in statistics. It is generally referred to as the **normal distribution**.
Mean = Median = Mode
- › Real-life examples of a symmetrical (or normal) distribution are
 - the heights of a random sample of people
 - the intelligence quotients (IQ) of a population.

2. Positive skew

- › When a distribution has most of the data at the lower values, we say it has a **positive skew**. The following histogram shows a positive skew as most of the data, represented by the higher bars, is mainly to the left.

If there is a positive skew, most of the data is to the left.

Notice that there is a long tail to the right of the distribution.

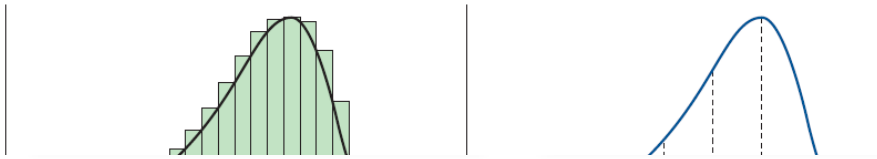


- › Real-life examples of a distribution with a positive skew are
 - the number of children in a family
 - the age at which people first learn to ride a bicycle
 - the age at which people marry.

3. Negative skew

› When a distribution has most of the data at the higher values, we say that the distribution has a **negative skew**.

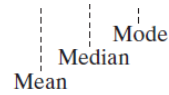
When a distribution has a negative skew, the tail will be to the left.



- › Real-life examples of a distribution with a negative skew are
 - (i) the ages at which people have to get their first pair of reading glasses
 - (ii) the heights of players playing in a professional basketball league.

Tail to the left

In a distribution with a **positive** skew, the tail is to the **right**; with a **negative** skew, the tail is to the **left**.

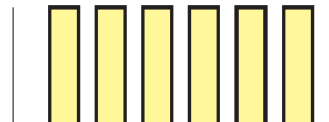


$\text{Mean} < \text{Median} < \text{Mode}$

4. Uniform distributions

In a uniform distribution, the data is evenly spread throughout.

It does not have a modal class.



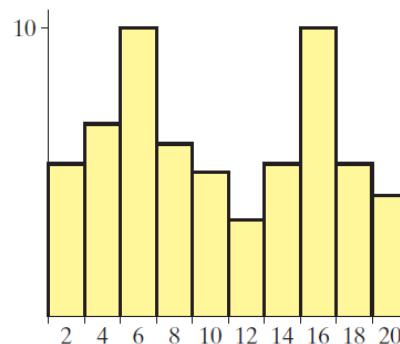
5. Bimodal distributions

This distribution has two modes.

It is called a **bimodal distribution**.

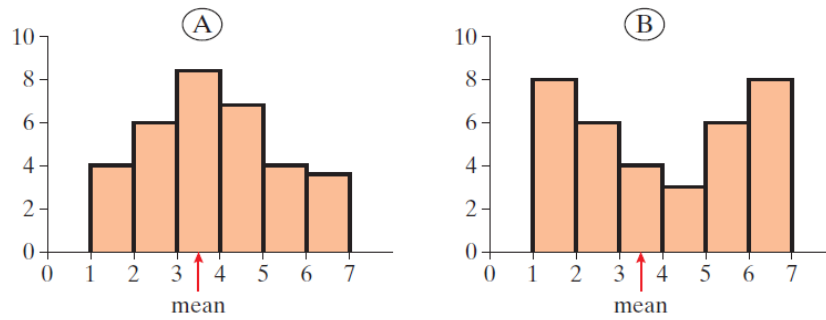
The modes are 6 and 16.

A distribution that has three or more modes is said to be **multimodal**.



6. Distributions and standard deviation

Consider the two distributions, (A) and (B), shown below:



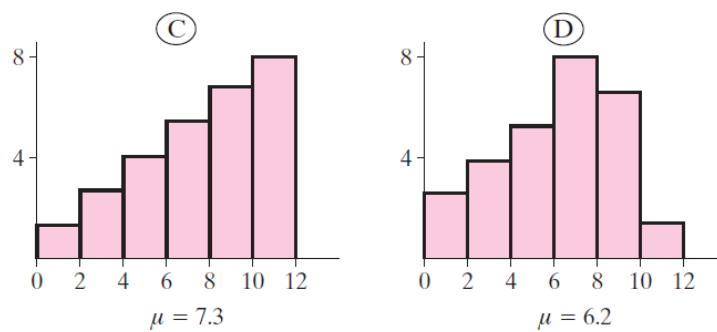
In distribution (A), most of the data lies between 2 and 5.

In distribution (B), the data is more spread out and further from the mean than the data in distribution (A).

The more spread out the data is in a distribution, the greater the standard deviation will be.

In the distributions above, we can conclude that (B) has a higher standard deviation than (A).

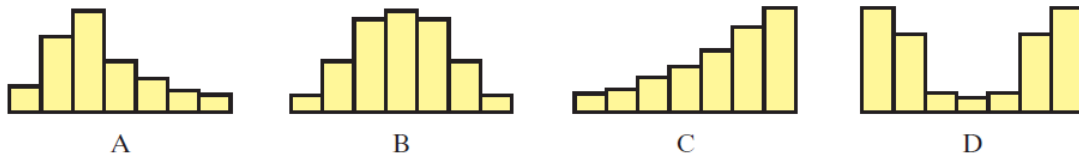
In the two distributions (C) and (D) below, the mean μ is given in each.



From the diagrams, we can see that more of the data is further from the mean in (C) than in (D).

Thus distribution (C) has the greater standard deviation.

10. The shapes of the histograms of four different sets of data are shown below.



(i) Complete the table below, indicating whether the statement is correct (✓) or incorrect (✗) with respect to each data set.

	A	B	C	D
The data is skewed to the left.			✓	
The data is skewed to the right.	✓			
The mean is equal to the median.		✓		✓
The mean is greater than the median.	✓			
There is a single mode.	✓	✓	✓	

(ii) Assume that the four histograms are drawn on the same scale. State which of them has the largest standard deviation, and justify your answer.

(D) because it has greatest spread of data from mean.