Introduction to the Normal distribution

# Normal random variables

The Normal distribution is by far the most important and useful probability distribution in statistics, with many applications in economics, engineering, astronomy, medicine, error and variation analysis, etc. The Normal distribution is often called the *bell curve*, due to its distinctive shape.



The Normal distribution is best for modelling continuous random variables that tend to clump around some central average value, and become more unlikely the further away they are. Some good examples of random variables that could be modelled with the Normal distribution are

* the weight of a new-born baby



* the blood pressure of an adult



* the height of adult English male criminals



## Activity

Below is a list of random variables. Decide which of them could be suitably modelled by a Normal distribution.

* the time required to fly from London to Paris
* the time until someone next sends you a text message
* the length of a hammer mass-produced by a machine
* the number of conservative councillors elected at the next local council election
* the return on an investment
* the time taken to download a file
* the distance until the next pot-hole when driving on the motorway
* the length of a carrot grown in your garden
* the number of television insides a 3-bedroom house
* the volume of milk produced by a fully-grown cow
* the battery life of a new phone from a particular manufacturer
* the number of even numbers that will appear in the lottery

# Modelling with the Normal distribution

When modelling a random variable with the Normal distribution, we tend to use very specific notation. Often, we choose a capital letter, like to represent the random variable itself. For example, we could let be the weight of a new-born baby in the UK. To declare that we intend to model with a Normal distribution, we write and say “ follows a Normal distribution with mean and variance ”. The values of and are normally specific values, so you are likely to see declarations like and . In the first example, and . In the second example, and .

What is the effect of and ?

The first value, , represents the mean value of the random variable and will dictate the centre-values of the bell curve. Below are the bell curves corresponding to , and . Make sure you are able to read these centre-values from bell curves.



The second value, , represents the variance of the random variable and will dictate the spread of the bell curve. Below are the bell curves corresponding to , and . It’s not possible to read the value of from the bell curve directly, but generally speaking as increases then the possible range of values modelling by the Normal distribution spreads out.



### Activity

Match the Normal distributions with their corresponding bell curves.



# Probabilities from the Normal distribution

Once a Normal distribution has been declared for a random variable we are able to use this distribution to find probabilities associated with the random variable. Suppose that is a random variable and we are to model it as . Then the probability that is between two values, and , will be given by the area under the bell curve between and . We write this as .

For example, if .



Then the probability that is between and is the area under the bell curve between and .



In this example, it turns out that , which is 82%. We won’t yet be considering how to calculate the probability. For now, it’s enough to make the connection between the probability, the limits (2 and 5) and the area.

It’s also possible to ask for one-sided probabilities, such as



or



We can also ask for slightly more complicated probabilities, like



and



### Activity

Suppose that , and .

**

Match each of the probabilities with the correct image

1. 
2. 
3. 
4. 
5. 
6. 
7. 
8. 
9. 
10. 
11. 