Statistics for Economics,

Ch03
Probability theory and statistical inference

3.1 Probability theory; the building blocks:

- to establish a vocabulary that we will subsequently use.
- ● An experiment is an action such as flipping a coin, which has a number of possible outcomes or events, such as heads or tails.
- ● A trial is a single performance of the experiment, with a single outcome.
- ● The sample space consists of all the possible outcomes of the experiment. The outcomes for a single toss of a coin are {heads, tails}, for example, and these constitute the sample space for a toss of a coin. The outcomes in the sample space are mutually exclusive, which means that the occurrence of one rules out all the others. One cannot have both heads and tails in a single toss of a coin. As a further example, if a single card is drawn at random from a pack, then the sample space may be drawn as in Figure below. Each point represents one card in the pack and there are 52 points altogether. (The sample space could be set out in alternative ways. For instance, one could write a list of all the cards: ace of spades, king of spades, . . . , two of clubs. One can choose the representation most suitable for the problem at hand.)
- ● With each outcome in the sample space we can associate a probability, which is the chance of that outcome occurring. The probability of heads is one-half; the probability of drawing the ace of spades from a pack of cards is one in 52, etc.

3.1 Probability theory; the building blocks:

	A	K	Q	J	10	9	8	7	6	5	4	3	2
♠	●	●	●	●	●	●	●	●	●	●	●	●	●
♥	●	●	●	●	●	●	●	●	●	●	●	●	●
♦	●	●	●	●	●	●	●	●	●	●	●	●	●
♣	●	●	●	●	●	●	●	●	●	●	●	●	●

- There are restrictions upon the probabilities we can associate with the outcomes in the sample space.
- These are needed to ensure that we do not come up with self-contradictory results; for example, it would be odd to arrive at the conclusion that we could expect heads more than half the time and tails more than half the time.
- To ensure our results are always consistent, the following rules apply to probabilities:
- 1. The probability of an event must lie between 0 and 1, i.e.

$$0 \leq \Pr(A) \leq 1, for\ any\ event\ A \qquad (1)$$

3.1 Probability theory; the building blocks:

- The explanation is straightforward. If A is certain to occur it occurs in 100% of all trials and so its probability is 1. If A is certain not to occur then its probability is 0, since it never happens however many trials there are. As one cannot be more certain than certain, probabilities of less than 0 or more than 1 can never occur, and equation (1) follows.
- 2. The sum of the probabilities associated with all the outcomes in the sample space is 1. Formally

$$\Sigma P_i = 1 \tag{2}$$

- where P_i is the probability of event i occurring. This follows from the fact that one, and only one, of the outcomes must occur, since they are mutually **exclusive** and also exhaustive, i.e. they define all the possibilities.
- 3. Following on from equation (2.2) we may define the **complement** of an event as everything in the sample space apart from that event. The complement of heads is tails, for example.
- If we write the complement of A as not-A then it follows that $\Pr(A) + \Pr(not - A) = 1$ and hence

$$\Pr(not - A) = 1 - \Pr(A) \tag{3}$$

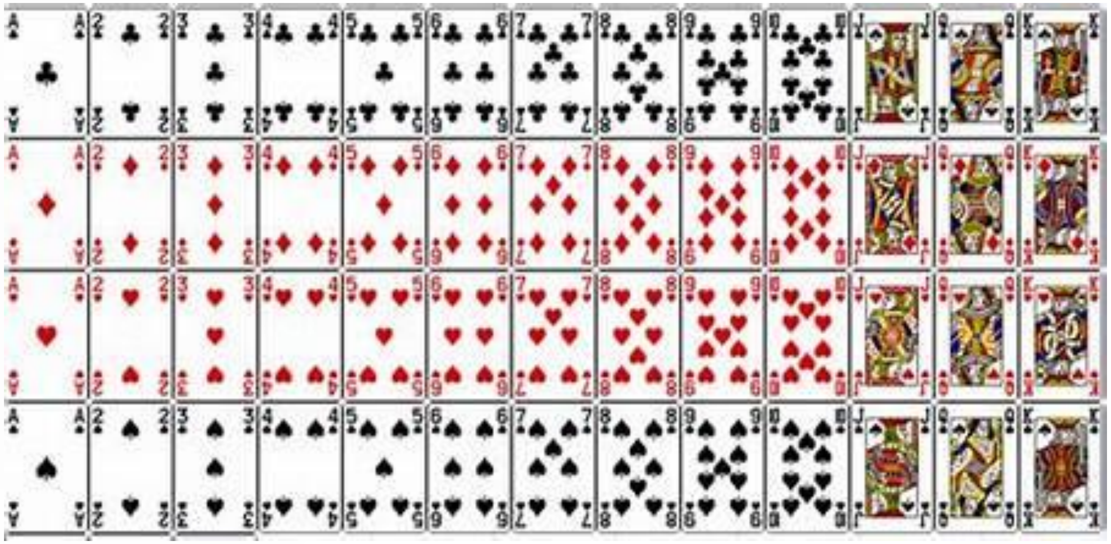
3.1 Probability theory; the building blocks:

- *Compound events*

- Most practical problems require the calculation of the probability of a set of outcomes rather than just a single one, or the probability of a series of outcomes in separate trials.

- For example, the probability of drawing a spade at random from a pack of cards encompasses 13 points in the sample space (one for each spade).

- This probability is 13 out of 52, or one-quarter, which is fairly obvious; but for more complex problems the answer is not immediately evident. We refer to such sets of outcomes as **compound events**.



3.1 Probability theory; the building blocks:

- Some examples are getting a five or a six on a throw of a die or drawing an ace and a queen to complete a ‘straight’ in a game of poker.
- It is sometimes possible to calculate the probability of a compound event by examining the sample space, as in the case of drawing a spade above. However, in many cases this is not so, for the sample space is too complex or even impossible to write down.
- For example, the sample space for three draws of a card from a pack consists of over 140000 points! (A typical point might be, for example, the ten of spades, eight of hearts and three of diamonds.)
- An alternative method is needed. Fortunately there are a few simple rules for manipulating probabilities which help us to calculate the probabilities of compound events.
- If the previous examples are examined closely it can be seen that outcomes are being compounded using the words ‘or’ and ‘and’: ‘. . . five or six on a single throw . . .’; ‘. . . an ace and a queen . . .’. ‘And’ and ‘or’ act as operators, and compound events are made up of simple events compounded by these two operators. The following rules for manipulating probabilities show how to use these operators and thus how to calculate the probability of a compound event.

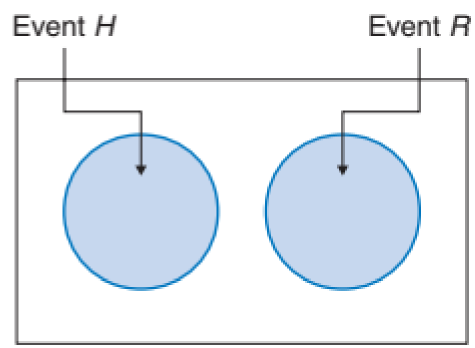
3.1 Probability theory; the building blocks:

- ***The addition rule***

- The addition rule is associated with 'or'. When we want the probability of one event or another occurring, we add the probabilities of each. More formally, the probability of A or B occurring is given by

$$\Pr(A\ or\ B) = \Pr(A) + \Pr(B)$$

- We illustrate this in the next Figure for the motor racing example. The circle labelled 'Event H' represents Hamilton winning, 'Event R' is Rosberg winning. The rest of the sample space represents a win by a different driver.



- Note that the two circles do not overlap, since the events are **mutually exclusive**; if Hamilton wins, then Rosberg cannot also win and vice versa.

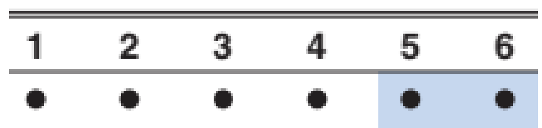
- An outcome in H or in R will mean a Mercedes victory, so we simply add the probabilities of each event to get the answer desired.

3.1 Probability theory; the building blocks:

- To reinforce the idea, consider getting a five or a six on a roll of a die1. This is

$$\Pr(6\ or\ 5) = \Pr(5) + \Pr(6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

- This answer can be verified from the sample space, as shown in Figure below. Each dot represents a simple event (one to six). The compound event is made up of two of the six points, shaded in Figure below, so the probability is 2/6 or 1/3.



- However, this is not a general solution to this type of problem, i.e. it does not always give the right answer, as can be seen from the following example. What is the probability of a queen or a spade in a single draw from a pack of cards?

- $\Pr(Q) = \frac{4}{52}$ (four queens in the pack) and $\Pr(A) = \frac{13}{52}$ (13 spades), so applying the above equation gives

$$\Pr(Q\ or\ S) = \Pr(Q) + \Pr(s) = \frac{4}{52} + \frac{13}{52} = \frac{17}{52}$$

3.1 Probability theory; the building blocks:

- The problem is that one point in the sample space (the one representing the queen of spades) is double-counted in the equation, <u>once as a queen and again as a spade</u>.

- The event 'drawing a queen and a spade' is possible, and gets **double counted**. This issue can again be illustrated using a Venn diagram below. It is worth noting that the two circles overlap.

	A	K	Q	J	10	9	8	7	6	5	4	3	2
♠	●	●	●	●	●	●	●	●	●	●	●	●	●
♥	●	●	●	●	●	●	●	●	●	●	●	●	●
♦	●	●	●	●	●	●	●	●	●	●	●	●	●
♣	●	●	●	●	●	●	●	●	●	●	●	●	●

- The overlap area is called the <u>intersection</u> of the two sets Q and S and represents any card that is both a queen and a spade (i.e. just the queen of spades). The intersection of the sets is written $Q \cap S$.

- Therefore, we wish to consider all of the outcomes within the circles, counted once only. Formally this is known as the union of the two sets, written $Q \cup S$.

- if we count all the outcomes in Q and then all those in S, we will double count those in the intersection. Hence, we need to subtract the intersection.

3.1 Probability theory; the building blocks:

- In the language of sets we have:

$$Q \cup S = Q + S - (Q \cap S)$$

- This carries over to probabilities. Equation has to be modified by subtracting the probability of getting a queen and a spade, to eliminate this double counting.

- The general rule is therefore

$$\Pr(A \; or \; B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

- In general, therefore, one should use this formula when two events are NOT mutually exclusive.

3.1 Probability theory; the building blocks:

- ***The multiplication rule***
- The multiplication rule is associated with use of the word <u>'and'</u> to combine events.
- Consider the example of a mother with two children. What is the probability that they are both boys? This is really a compound event: that the first child is a boy and the second is also a boy. It corresponds to the intersection of the two sets in a Venn diagram. Assume that in any single birth a boy or girl is equally likely, so $\Pr(boy) = \Pr(girl) = 0.5$. Denote by $\Pr(B1)$ the probability of a boy on the first birth and by $\Pr(B2)$ the probability of a boy on the second.
- Thus the question asks for $\Pr(B1 \; and \; B2)$ and this is given by:

$$\Pr(B1 \; and \; B2) = \Pr(B1) \times \Pr(B2) = 0.5 * 0.5 = 0.25$$

- Intuitively, the multiplication rule can be understood as follows. One-half of mothers have a boy on their first birth and of these, one-half will again have a boy on the second. Therefore, a quarter (a half of one-half) of mothers have two boys.

3.1 Probability theory; the building blocks:

The multiplication rule and independence

The conditional probability, more precisely the probability of B2 conditional upon B1. In words, it means the probability of having a second boy after the first is a boy. Let us drop the independence assumption and suppose the following

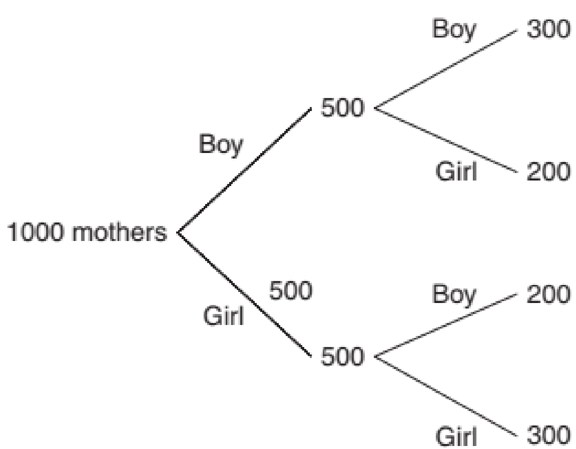
$$\Pr(B1) = \Pr(G1) = 0.5$$

i.e. boys and girls are equally likely on the first birth (as previously assumed), but

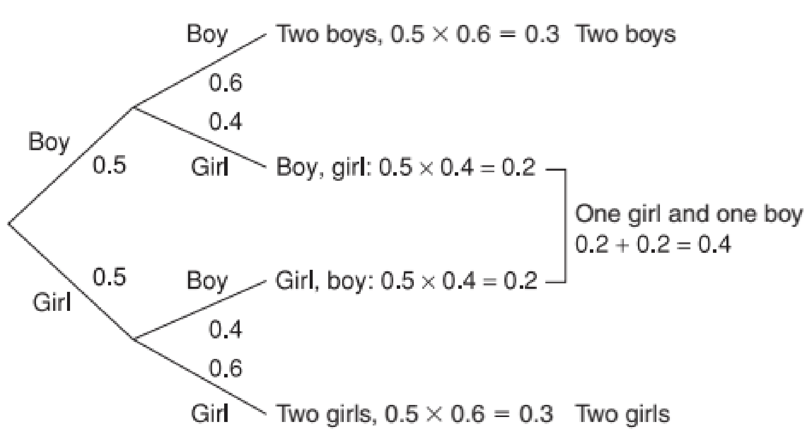
$$\Pr(B2|B1) = \Pr(G2|G1) = 0.6$$

i.e. a boy is more likely to be followed by another boy, and a girl by another girl.

This new situation can be usefully illustrated with a tree diagram, either using frequencies or probabilities.



or probabilities



3.1 Probability theory; the building blocks:

- these figures illustrate what is the probability of two boys, two girls and one boy, one girl
- Thus in general we have:

$$\Pr(A \text{ } and \text{ } B) = \Pr(A) \times \Pr(B|A)$$

- Independence may therefore be defined as follows: two events, A and B, are independent if the probability of one occurring is not influenced by the fact of the other having occurred.
- Formally, if A and B are independent, then

$$\Pr(B|A) = P(A|not \text{ } A) = \Pr(B)$$

- The concept of independence is an important one in statistics, as it usually simplifies problems considerably. If two variables are known to be independent, then we can analyse the behaviour of one without worrying about what is happening to the other variable. For example, sales of computers are independent of temperature, so if one is trying to predict sales next month, one does not need to worry about the weather. In contrast, ice cream sales do depend on the weather, so predicting sales accurately requires one to forecast the weather first.

3.1 Probability theory; the building blocks:

- ***Combining the addition and multiplication rules***
- More complex problems can be solved by suitable combinations of the addition and multiplication formula.
- For example, what is the probability of a mother having one child of each sex? This could occur in one of two ways: a girl followed by a boy or a boy followed by a girl.
- It is important to note that these are two different routes to the same outcome.
- $\Pr(1\ girl, 1\ boy) = \Pr\big((G1\ and\ B2) or\ (B1\ and\ G2)\big)$
- $= \Pr(G1) * \Pr(B2 \mid G1) + \Pr(B1) * \Pr(G2 \mid B1)$
- $= (0.5 * 0.4) + (0.5 * 0.4)$
- $= 0.4$

3.1 Probability theory; the building blocks:

- ***Combinations and permutations***
- The methods described above are adequate for solving fairly simple probability puzzles but fall short when more complex questions are asked.
- For example, what is the probability of three girls and two boys in a family of five children?
- The tree diagram can obviously be extended to cover third and subsequent children, but the number of branches rapidly increases (in geometric progression).
- It takes time to draw the diagram (it has 32 end points), and identify the relevant paths and associated probabilities, and it is easy to make an error.
- If you do this correctly, you will find that there are 10 relevant paths through the diagram (e.g. GGGBB or GGBBG) and each individual path has a probability of 1/32 (½ raised to the power 5), so the answer is 10/32. Note that we are once again assuming independent events here, so the probability of having a boy or a girl is always 0.5.

3.1 Probability theory; the building blocks:

- Far better would be to use a formula in a complex case like this. To develop this, we introduce the ideas of combinations and permutations.
- The strategy for finding the probability of three girls and two boys in five children is two-fold:
- (1) Work out the probability of three girls and two boys in one particular order (e.g. GGGBB).

  This is $0.5^5 = \frac{1}{32}$.

- (2) Use a formula to work out the number of different orderings, in this case 10.
- It is this second point that we now focus on. How can we establish the number of ways of having three girls and two boys in a family of five children? One way would be to write down all the possible orderings:
- GGGBB GGBGB GGBBG GBGGB GBGBG
- GBBGG BGGGB BGGBG BGBGG BBGGG

3.1 Probability theory; the building blocks:

- This shows that there are 10 such orderings, so the probability of three girls and two boys in a family of five children is 10/32. In more complex problems, this soon becomes difficult or impossible. The record number of children born to a British mother is 39 (!) of whom 32 were girls. The appropriate tree diagram has over five thousand billion paths through it, and drawing one line (i.e. for one child) per second would imply 17 433 years to complete the task.

- rather than do this, we use the combinatorial formula to find the answer.

- Suppose there are n children, r of them girls, then the number of orderings, denoted nCr, is obtained from

$$nCr = \frac{n!}{r!\,(n-1)!}$$

- **<u>E.g.</u>** If there were four girls out of five children, then the number of orderings or combinations would be …

- This gives five possible orderings, i.e. the single boy could be the first, second, third, fourth or fifth born.

3.2 Bayes' theorem:

- Bayes' theorem is a factual statement about probabilities which in itself is uncontroversial.
- However, the use and interpretation of the result is at the heart of the difference between classical and Bayesian statistics. The theorem itself is easily derived from first principles.

$$\Pr(A\ and\ B) = \Pr(A|B) \times Pr(B)$$

- Hence, $\Pr(A|B) = \frac{\Pr(A \cup B)}{\Pr(B)}$
- Expanding both top and bottom of the right-hand side

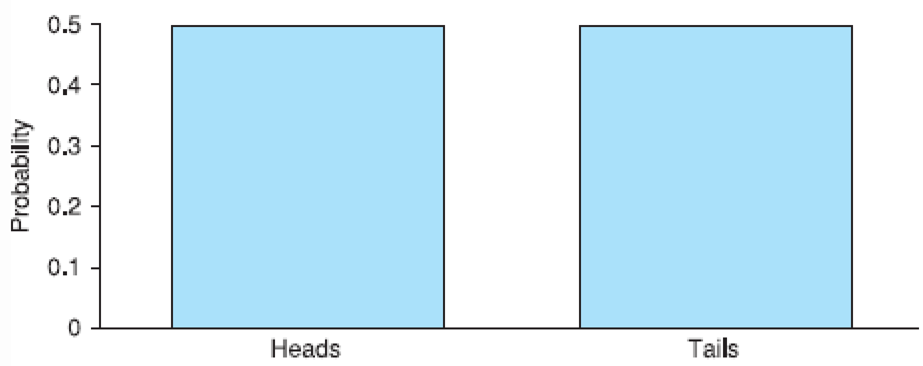
$$\Pr(A|B) = \frac{\Pr(A|B) \times Pr(B)}{\Pr(B|A) \times Pr(A) + \Pr(\mathrm{A|not\ A}) \times \Pr(not\ A)}$$

- This Equation is known as Bayes' theorem and is a statement about the probability of the event A, conditional upon B having occurred. The following example demonstrates its use.

Ch04
Probability distributions

4.1 Introduction:

- In this chapter the probability concepts introduced in Chapter 3 are generalized by using the idea of a probability distribution.
- A probability distribution lists, in some form, all the possible outcomes of a probability experiment and the probability associated with each one. For example, the simplest experiment is tossing a coin, for which the possible outcomes are heads or tails, each with probability one-half.
- The probability distribution can be expressed in a variety of ways: in words, or in a graphical or mathematical form. For tossing a coin, the graphical form is shown in Figure below,



- And the mathematical form is $\Pr(H) = 0.5$ and $\Pr(T) = 0.5$

4.1 Introduction:

- Some probability distributions occur often and so are well known. Because of this they have names so we can refer to them easily; for example, the Binomial distribution or the Normal distribution. In fact, each constitutes a family of distributions.

- A single toss of a coin gives rise to one member of the Binomial distribution family; two tosses would give rise to another member of that family.

- These two distributions differ in the number of tosses. If a biased coin were tossed, this would lead to yet another Binomial distribution, but it would differ from the previous two because of the different probability of heads.

- Members of the Binomial family of distributions are distinguished either by the number of tosses or by the probability of the event occurring. These are the two parameters of the distribution and tell us all we need to know about the distribution. Other distributions might have different numbers of parameters, with different meanings. Some distributions, for example, have only one parameter.

- In order to understand fully the idea of a probability distribution a new concept is first introduced, that of a random variable.

4.2 Random variables:

- As an example of the random variable: the result of the toss of a coin, or the number of boys in a family of five children.
- A random variable is one whose outcome or value is the result of chance and is therefore unpredictable, although the range of possible outcomes and the probability of each outcome may be known.
- It is impossible to know in advance the outcome of a toss of a coin for example, but it must be either heads or tails, each with probability one-half.
- The number of heads in 250 tosses is another random variable, which can take any value between zero and 250, although values near 125 are the most likely.
- You are very unlikely to get 250 heads from tossing a fair coin!
- Intuitively, most people would 'expect' to get 125 heads from 250 tosses of the coin, since heads comes up half the time on average.
- This suggests we could use the expected value notation introduced in before and write $E(X) = 125$, where X represents the number of heads obtained from 250 tosses. This usage is indeed valid and we will explore this further below. It is a very convenient shorthand notation.

4.3 The Binomial distribution:

- The Binomial distribution arises whenever the underlying probability experiment has just two possible outcomes, for example heads or tails from the toss of a coin.
- Even if the coin is tossed many times (so one could end up with one, two, three . . . , etc., heads in total) the underlying experiment has only two outcomes, so the Binomial distribution should be used.
- A counter-example would be the rolling of die, which has six possible outcomes (in this case the Multinomial distribution, not covered in this book, would be used).
- Note, however, that if we were interested only in rolling a six or not, we could use the Binomial by defining the two possible outcomes as ‘six’ and ‘not-six’.
- It is often the case in statistics that by suitable transformation of the data we can use different distributions to tackle the same problem.

4.3 The Binomial distribution:

- For the Binomial distribution to apply we first need to assume independence of successive events and we shall assume that, for any birth

$$\Pr(boy) = P = 0.5$$

- It follows that $\Pr(girl) = 1 - \Pr(boy) = 1 - P = 0.5$
- Although we have $P = 0.5$ in this example, the Binomial distribution can be applied for any value of P between 0 and 1.
- First we consider the case of $r = 5, n = 5$, (five boys in five births). This probability is found using the multiplication rule:

$$\Pr(r = 5) = P * P * P * P * P = P^5 = (0.5)^5 = \frac{1}{32}$$

- The probability of four boys (and then implicitly one girl) is $\Pr(r = 4) = P * P * P * P * (1 - P) = \frac{1}{32}$
- The formula for four boys and one girl is $\Pr(r = 4) = 5C4 * P^4 * (1 - P)$

For three boys (and two girls) we obtain $\Pr(r = 3) = 5C3 * P^3 * (1 - P)^2 = 10 * \frac{1}{8} * \frac{1}{4} = \frac{10}{32}$

4.3 The Binomial distribution:

- A fairly clear pattern emerges. The probability of r boys in n births is given by

$$\Pr(r) = nCr * P^r * (1 - P)^{n-r}$$

- this is known as the Binomial formula or distribution.
- The Binomial distribution is appropriate for analysing problems with the following characteristics:
  1. There is a number (n) of trials.
  2. Each trial has only two possible outcomes, 'success' (with probability P) and 'failure' (probability 1 - P) and the outcomes are independent between trials.
  3. The probability P does not change between trials.
- Since the Binomial distribution depends only upon the two values n and P, we can use a shorthand notation rather than the formula itself.
- A random variable r, which has a Binomial distribution with the parameters n and P, can be written in general terms as

$$r \sim B(n, p)$$

4.3 The Binomial distribution:

- ***The mean and variance of the Binomial distribution***
- The mean and variance are most easily calculated by drawing up a relative frequency table based on the Binomial frequencies.
- The mean of this distribution is given by

$$E(r) = \frac{\sum r \times \Pr(r)}{\sum \Pr(r)}$$

- The variance is given by

$$V(r) = \frac{\sum r^2 \times \Pr(r)}{\sum \Pr(r)} - \mu^2$$

4.3 The Binomial distribution:

- *Example:* If a die is thrown four times, what is the probability of getting two or more sixes?
- This is a problem involving repeated experiments (rolling the die) with but two types of outcome for each roll: success (a six) or failure (anything but a six).
- Note that we combine several possibilities (scores of 1, 2, 3, 4 or 5) together and represent them all as failure.
- The probability of success (one-sixth) does not vary from one experiment to another, and so use of the Binomial distribution is appropriate.
- The values of the parameters are $n = 4$ and $P = \frac{1}{6}$.
- Denoting by r the random variable 'the number of sixes in four rolls of the die', then

$$r \sim B(4,6)$$

- The probabilities of two, three and four sixes are then given by…
- Since these events are mutually exclusive, the probabilities can simply be added together to get the desired result, which is 0.132, or 13.2%. This is the probability of two or more sixes in four rolls of a die.

4.3 The Binomial distribution:

- **Exercise 1**
- (a) If the probability of a randomly drawn individual having blue eyes is 0.6, what is the probability that four people drawn at random all have blue eyes?
- (b) What is the probability that two of the sample of four have blue eyes?
- (c) For this particular example, write down the Binomial formula for the probability of r blue eyed individuals, $r = 0 \ldots 4$. Confirm that the calculated probabilities sum to one.
- (d) Calculate the mean and variance of the number of blue-eyed individuals.
- (e) Draw a graph of this Binomial distribution and on it mark the mean value and the mean value $+/-$ one standard deviation.

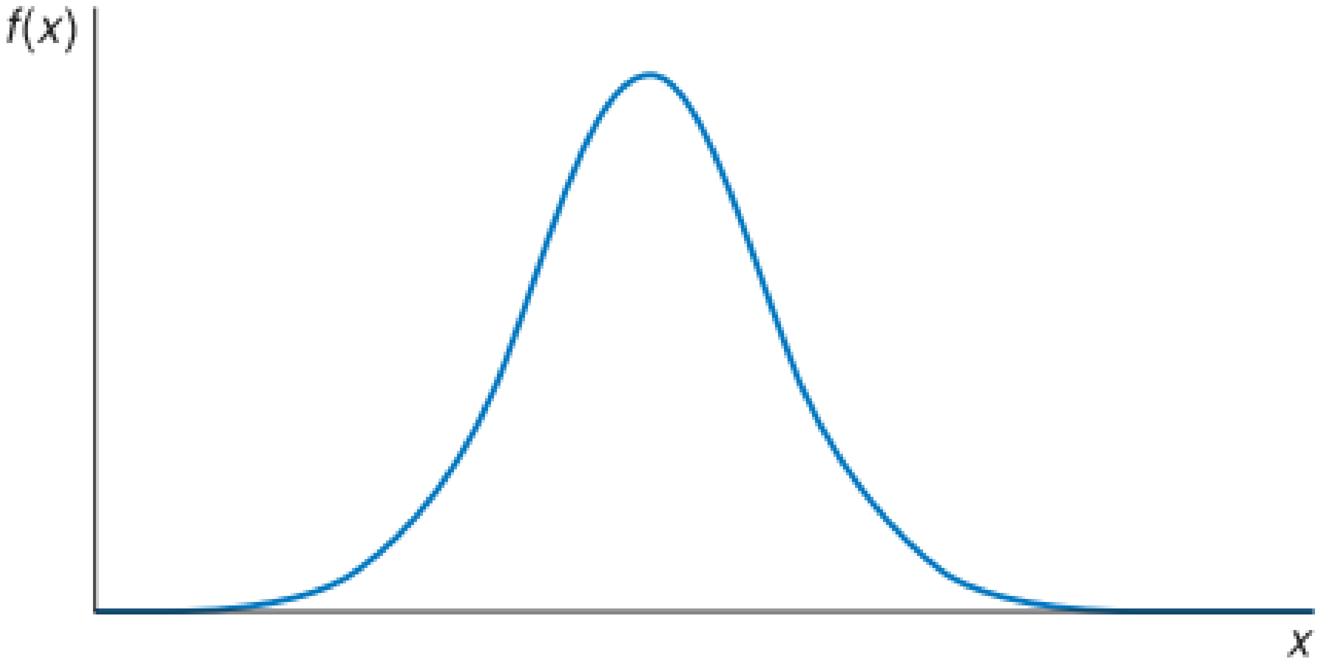
4.4 The Normal distribution:

- The Binomial distribution applies when there are two possible outcomes to an experiment.
- For instance, the (random) arrival time of a train is a continuous variable and cannot be analysed using the Binomial.
- There are many probability distributions in statistics, developed to analyse different types of problem.
- Many random variables turn out to be Normally distributed.
- Men's (or women's) heights are Normally distributed. IQ (the measure of intelligence) is also Normally distributed. Another example is of a machine producing (say) bolts with a nominal length of 5 cm which will actually produce bolts of slightly varying length (these differences would probably be extremely small) due to factors such as wear in the machinery, slight variations in the pressure of the lubricant, etc.
- These would result in bolts whose length varies, in accordance with the Normal distribution.
- This sort of process is extremely common, with the result that the Normal distribution often occurs in everyday situations.

4.4 The Normal distribution:

- Unlike the Binomial, the Normal distribution applies to continuous random variables such as height, and a typical Normal distribution is illustrated in Figure below.
- Since the Normal distribution is a continuous one, it can be evaluated for any values of x, not just for integers as was the case for the Binomial distribution. The figure illustrates the main features of the distribution:
- It is unimodal, having a single, central peak. If this were men's heights, it would illustrate the fact that most men are clustered around the average height, with a few very tall and a few very short people.
- It is symmetric, the left and right halves being mirror images of each other.
- It is bell-shaped.
- It extends continuously over all the values of x from minus infinity to plus infinity, although the value of $f(x)$ becomes extremely small as these values are approached (the presentation of this figure being of only finite width, this last characteristic is not faithfully reproduced).
- This also demonstrates that most empirical distributions (such as men's heights) can only be an approximation to the theoretical ideal, although the approximation is close and good enough for practical purposes.

4.4 The Normal distribution:



4.4 The Normal distribution:

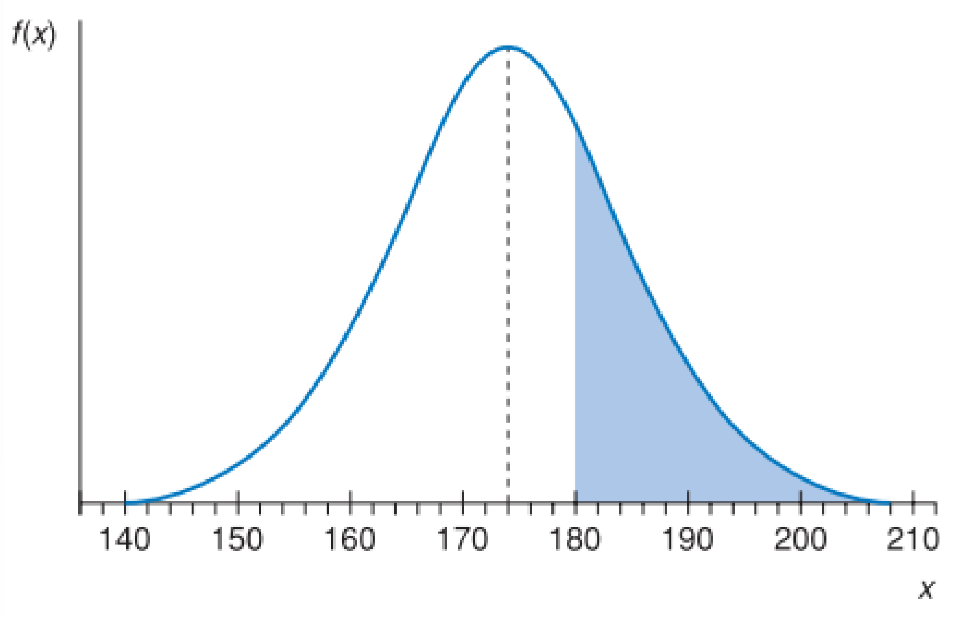
- Note that we have labelled the y-axis '$f\ (x)$' rather than '$\Pr(x)$' as we did for the Binomial distribution.
- This is because it is areas under the curve that represent probabilities, not the heights.
- With the Binomial, which is a discrete distribution, one can legitimately represent probabilities by the heights of the bars. For the Normal, although $f(x)$ does not give the probability per se, it does give an indication: you are more likely to encounter values from the middle of the distribution (where f (x) is greater) than from the extremes.
- In mathematical terms, the formula for the Normal distribution is (x is the random variable)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma}}$$

- Like the Binomial, the Normal is a family of distributions differing from one another only in the values of the parameters μ and σ.
- The Normal is another two-parameter family of distributions like the Binomial, and once the mean μ and the standard deviation σ are known, the whole of the distribution can be drawn.
- The shorthand notation for a Normal distribution is $x \sim N(\mu, \sigma^2)$

4.4 The Normal distribution:

- What is the probability that a randomly selected man is taller than 180 cm?
- If all men are equally likely to be selected, this is equivalent to asking what proportion of men are over 180 cm in height.
- This is given by the area under the Normal distribution, to the right of $x = 180$, as in Figure below. The further from the mean of 174, the smaller the area in the tail of the distribution. One way to find this area would be to use equation above, but this requires the use of sophisticated mathematics.



4.4 The Normal distribution:

- Since this is a frequently encountered problem, the answers have been set out in the tables of the standard Normal distribution. We can simply look up the solution.
- However, since there is an infinite number of Normal distributions (one for every combination μ and σ^2), it would be impossible to tabulate them all.
- The standard Normal distribution, which has a mean of zero and variance of one, is therefore used to represent all Normal distributions.
- Before the table can be consulted, therefore, the data have to be transformed so that they accord with the standard Normal distribution. The required transformation is the z score.
- This measures the distance between the value of interest (180) and the mean, measured in terms of standard deviations.
- Therefore, we calculate

$$z = \frac{x - \mu}{\sigma}$$

- Here $z \sim N(0,1)$

4.4 The Normal distribution:

- This transformation shifts the original distribution m units to the left and then adjusts the dispersion by dividing through by σ, resulting in a mean of 0 and variance 1.
- z is Normally distributed because x is Normally distributed.
- If x followed some other distribution, then z would not be Normal either.
- It is easy to verify the mean and variance of z using the rules for E and V operators encountered before.

$$E(z) = E\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma}[E(x-\mu)] = 0\,;\ rememebr\ E(x) = \mu$$

$$V(z) = V\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\left[V(x)\frac{\sigma^2}{\sigma^2}\right] = 1$$

- Back to our men example; $z = 0.63$
- This shows that 180 is 0.63 standard deviations above the mean, 174, of the distribution. This is a measure of how far 180 is from 174 and allows us to look up the answer in tables.

4.4 The Normal distribution:

- The task now is to find the area under the standard Normal distribution to the right of 0.63 standard deviations above the mean.
- This answer can be read off directly from the table of the standard Normal distribution, included as Table A2 in the Appendix.

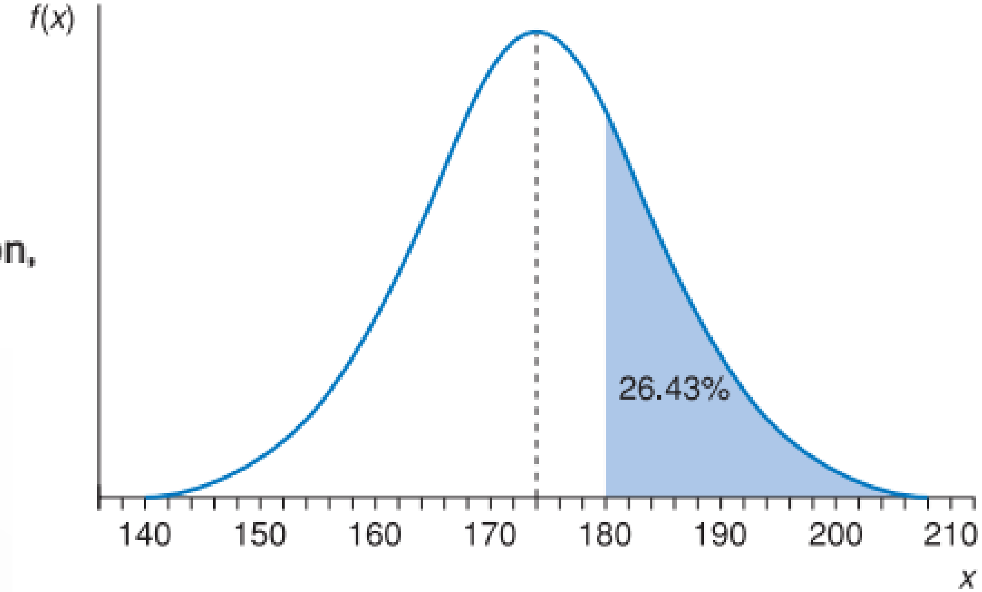
z	0.00	0.01	0.02	0.03	...	0.09
0.0	0.5000	0.4960	0.4920	0.4880	...	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	...	0.4247
⋮	⋮	⋮	⋮	⋮	...	⋮
0.5	0.3085	0.3050	0.3015	0.2981	...	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	...	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	...	0.2148

4.4 The Normal distribution:

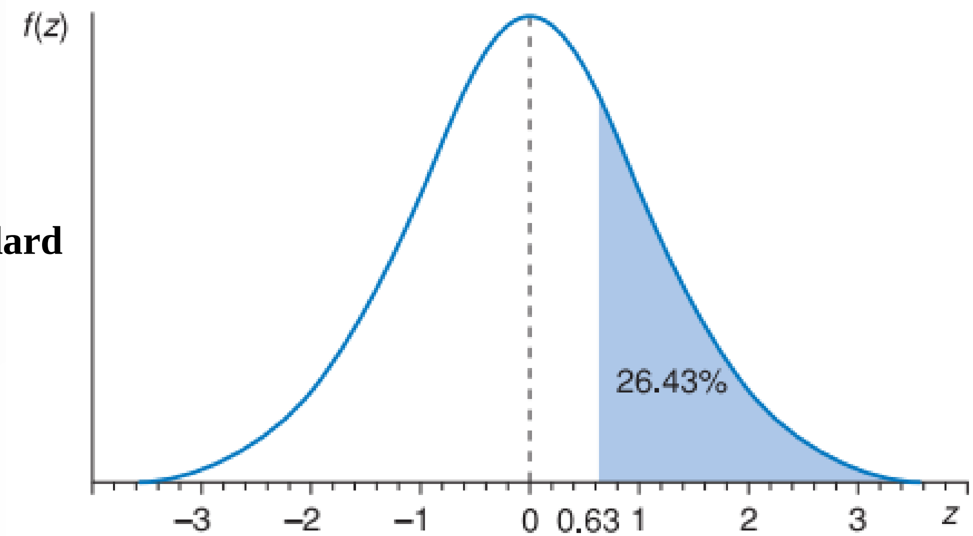
- The left-hand column gives the z score to one place of decimals.
- The appropriate row of the table to consult is the one for $z = 0.6$, which is shaded.
- For the second place of decimals (0.03) we consult the appropriate column, also shaded.
- At their intersection we find the value 0.2643, which is the desired area and therefore probability.
- In other words, 26.43% of the distribution lies to the right of 0.63 standard deviations above the mean. Therefore 26.43% of men are over 180 cm in height.
- Use of the standard Normal table is possible because, although there is an infinite number of Normal distributions, they are all fundamentally the same, so that the area to the right of 0.63 standard deviations above the mean is the same for all of them.
- As long as we measure the distance in terms of standard deviations, then we can use the standard Normal table.
- The process of standardization turns all Normal distributions into a standard Normal distribution with a mean of zero and a variance of one. This process is illustrated in Figure below.

4.4 The Normal distribution:

The Normal distribution, $\mu = 174, \sigma = 9.6$



The corresponding standard normal distribution

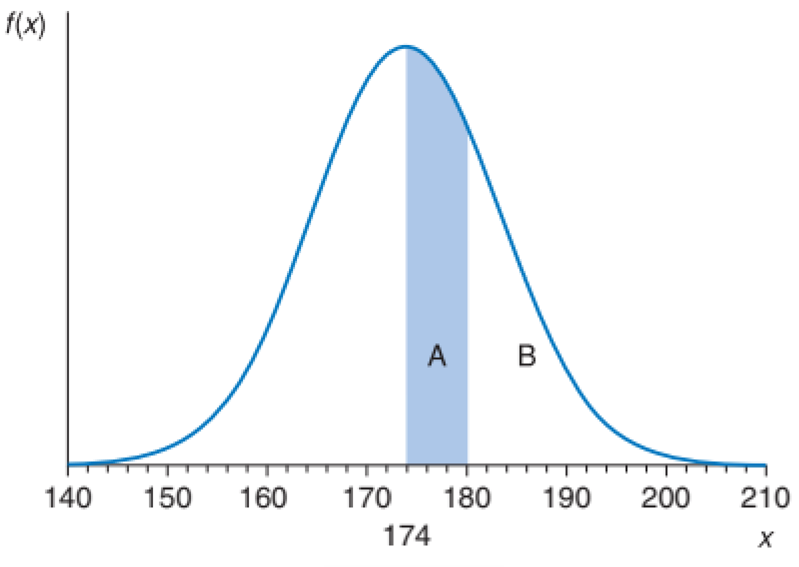


4.4 The Normal distribution:

- *Example:*
- Packets of cereal have a nominal weight of 750 grams, but there is some variation around this as the machines filling the packets are imperfect. Let us assume that the weights follow a Normal distribution. Suppose that the standard deviation around the mean of 750 is 5 grams.
- What proportion of packets weigh more than 760 grams?
- Solution:
- To summaries the problem we have: $z \sim N(750, 25)$ and we wish to find $\Pr(x > 760)$ …

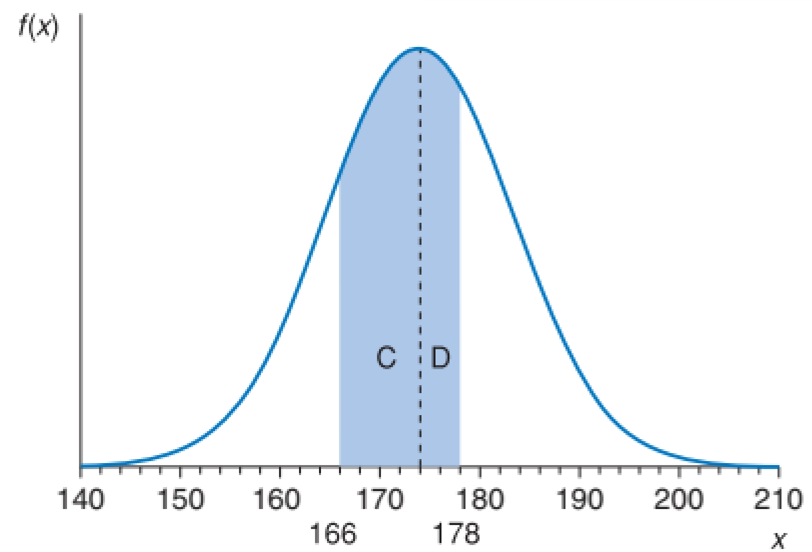
4.4 The Normal distribution:

- Since a great deal of use is made of the standard Normal tables, it is worth working through a couple more examples to reinforce the method.
- We have so far calculated that $\Pr(z > 0.63) = 0.2643$.
- Since the total area under the graph equals one (i.e. the sum of probabilities must be one), the area to the left of z = 0.63 must equal 0.7357, i.e. 73.57% of men are under 180 cm.
- It is fairly easy to manipulate areas under the graph to arrive at any required area.
- For example, what proportion of men are between 174 and 180 cm in height?
- It is helpful to refer to the next Figure at this point.



4.4 The Normal distribution:

- The size of area A is required. Area B has already been calculated as 0.2643.
- Since the distribution is symmetric, the area $A + B$ must equal 0.5, since 174 is at the centre (mean) of the distribution.
- Area A is therefore $0.5 - 0.2643 = 0.2357$.
- Therefore, 23.57% is the desired result.
- As a final exercise consider the question of what proportion of men are between **166 and 178 cm tall**.
- As shown in the next Figure, area C + D is wanted.



4.4 The Normal distribution:

- The only way to find this is to calculate the two areas separately and then add them together.
- For area D the z score associated with 178 is: $z = \frac{178-174}{9.6} = 0.42$
- As in z-Table, it indicates that the area in the right-hand tail, beyond $z = 0.42$, is 0.3372, so area

$D = 0.5 - 0.3372 = 0.1628$.

- For C, the z score is $z = \frac{166-174}{9.6} = -0.83$
- The minus sign indicates that it is the left-hand tail of the distribution, below the mean, which is being considered.
- Since the distribution is symmetric, it is the same as if it were the right-hand tail, so the minus sign may be ignored when consulting the table.
- Looking up $z = 0.83$ in the Table gives an area of 0.2033 in the tail, so area C is therefore $0.5 - 0.2033 = 0.2967$.
- Adding areas C and D gives $0.2967 + 0.1628 = 0.4595$. So nearly half of all men are between 166 and 178 cm in height.

4.5 The distribution of the sample mean:

- One of the most important concepts in statistical inference is the probability distribution of the mean of a random sample, since we often use the sample mean to tell us something about an associated population.
- Suppose that, from the population of adult males, a random sample of size $n = 36$ is taken, their heights measured and the mean height of the sample calculated.
- hat can we infer from this about the true average height of the population?
- To do this, we need to know about the statistical properties of the sample mean.
- The sample mean is a random variable because of the chance element of random sampling (different samples would yield different values of the sample mean).
- Since the sample mean is a random variable, it must have associated with it a probability distribution.
- We also refer to this as the sampling distribution of the sample mean, since the randomness is due to sampling.
- We therefore need to know, first, what is the appropriate distribution and, second, what are its parameters.

4.5 The distribution of the sample mean:

- From the definition of the sample mean we have

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \ \dots + x_n)$$

- where each observation, xi, is itself a Normally distributed random variable, with $x_i \sim N(\mu, \sigma^2)$, because each comes from the parent distribution with such characteristics.

- We stated earlier that men's heights are Normally distributed. We now make use of the following theorem to demonstrate first that x is Normally distributed:

- ***Theorem:*** Any linear combination of independent, Normally distributed random variables is itself Normally distributed.

- A linear combination of two variables x_1 and x_2 is of the form $w_1x_1 + w_2x_2$ where w_1 and w_2 are constants. This can be generalized to any number of x values.

- It is clear that the sample mean satisfies these conditions and is a linear combination of the individual x values (with the weight on each observation equal to $\frac{1}{n}$).

- As long as the observations are independently drawn, therefore, the sample mean is Normally distributed.

4.5 The distribution of the sample mean:

- We now need the parameters (mean and variance) of the distribution. For this we use the E and V operators once again:

$$E(\bar{x}) = \frac{1}{n}[E(x_1) + E(x_2) + \cdots E(x_n)] = \frac{1}{n}[\mu + \mu + \cdots + \mu] = \frac{n\mu}{n} = \mu$$

$$V(\bar{x}) = V\left(\frac{1}{n}[x_1 + x_2 + \cdots + x_n]\right) = \frac{1}{n}(V(x_1) + V(x_2) + \cdots + V(x_n)) = \frac{1}{n^2}(\sigma^2 + \sigma^2 + \cdots + \sigma^2) = \frac{\sigma^2}{n}$$

- Putting all this together, we have

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- ***Theorem:*** The mean, $\bar{x}$ of a random sample drawn from a population which has a Normal distribution with mean μ and variance σ^2, has a sampling distribution which is Normal, with mean μ and variance $\frac{\sigma^2}{n}$, where n is the sample size.

4.5 The distribution of the sample mean:

- This theorem can be used to solve a range of statistical problems. For example, what is the probability that a random sample of nine men will have a mean height greater than 180 cm? The height of all men is known to be Normally distributed with mean $\mu = 174$ cm and variance $\sigma^2 = 92.16$.

- The theorem can be used to derive the probability distribution of the sample mean. For the population we have:

$$x \sim N(174, 92.16) \text{ Hence the sample mean is } \bar{x}\ N\left(174, \frac{92.16}{9}\right)$$

- As for the question posed, the area to the right of 180 has to be found.

- This should by now be a familiar procedure. First the z score is calculated:

$$z = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{180 - 174}{\sqrt{\frac{92.16}{9}}} = 1.8$$

- Note that the z score formula is subtly different because we are dealing with the sample mean $\bar{x}$ rather than x itself. In the numerator we use $\bar{x}$ rather than x and in the denominator we use $\frac{\sigma^2}{n}$, not σ^2. This is $\bar{x}$ because has a variance $\frac{\sigma^2}{n}$, not σ^2, which is the population variance.

4.5 The distribution of the sample mean:

- $\sqrt{\frac{\sigma^2}{n}}$ is known as the standard error, to distinguish it from s, the standard deviation of the population. The principle behind the z score is the same, however: it measures how far a sample mean of 180 is from the population mean of 174, measured in terms of standard deviations.
- Looking up the value of $z = 1.88$ in the Table gives an area of 0.0311 in the right-hand tail of the Normal distribution.
- Thus, 3.11% of sample means will be greater than or equal to 180 cm when the sample size is **nine**. The desired probability is therefore 3.11%.
- Since this probability is quite small, we might consider the reasons for this.
- There are two possibilities:
- (a) through bad luck, the sample collected is not very representative of the population as a whole, or
- (b) the sample is representative of the population, but the population mean is not 174 cm after all.

4.5 The distribution of the sample mean:

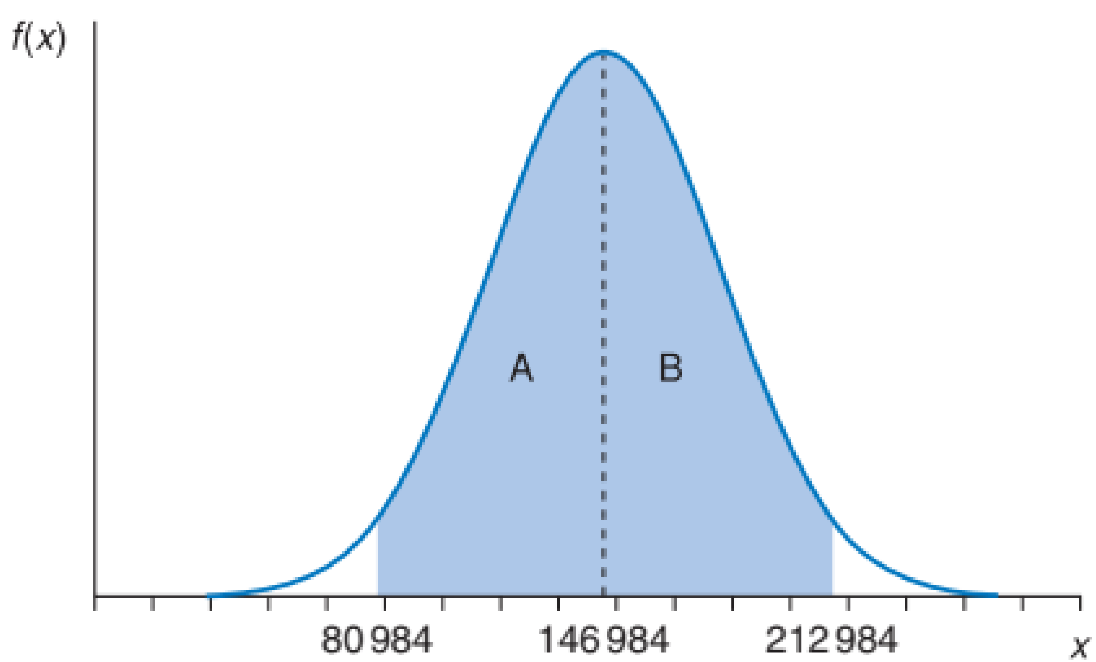
- ***Sampling from a non-Normal population***
- The previous theorem and examples relied upon the fact that the population followed a Normal distribution. But what happens if it is not Normal?
- After all, it is not known for certain that the heights of all adult males are exactly Normally distributed, and there are many populations which are not Normal (e.g. wealth, as shown in Chapter 1). What can be done in these circumstances?
- The answer is to use another theorem about the distribution of sample means (presented without proof).
- This is known as the Central Limit Theorem:
- ***Theorem:*** The mean, $\bar{x}$, of a random sample drawn from a population with mean μ and variance σ^2, has a sampling distribution which approaches a Normal distribution with mean μ and variance $\frac{\sigma^2}{n}$, as the sample size n approaches infinity.

4.5 The distribution of the sample mean:

- *Example:* As an example of the use of the Central Limit Theorem, we return to the wealth data of Chapter 1. Recall that the mean level of wealth was 186.875 (measured in £000) and the variance 80 306. Suppose that a sample of n = 50 people were drawn from this population. What is the probability that the sample mean is greater than 200 (i.e. £200 000)?
- Solution:
- On this occasion we know that the parent distribution is highly skewed, so it is fortunate that we have 50 observations. This should be ample for us to justify applying the Central Limit Theorem.
- The distribution of $\bar{x}$ is therefore $\bar{x} \sim N(186.875, \frac{80306}{50})$
- To find the area beyond a sample mean of 200, the z score is first calculated: $z = \frac{200-186.875}{\sqrt{\frac{80306}{50}}}$
- Referring to the standard Normal tables, the area in the tail is then found to be 37.07%. This is the desired probability. So there is a probability of 37.07% of finding a mean of £200 000 or greater with a sample of size 50. This demonstrates that there is quite a high probability of getting a sample mean which is relatively far from £186 875. This is a consequence of the high degree of dispersion in the distribution of wealth.

4.5 The distribution of the sample mean:

- Extending this example, we can ask: what is the probability of the sample mean lying within, say, £78 000 either side of the true mean of £186 875 (i.e. between £108 875 and £264 875)?
- Figure below illustrates the situation, with the desired area shaded.
- By symmetry, areas A and B must be equal, so we only need find one of them.



4.5 The distribution of the sample mean:

- For B, we calculate the z score: $z = \frac{264.875-186.875}{\sqrt{\frac{80306}{50}}}$

- From the standard Normal table, this cuts off approximately 2.5% in the upper tail,4 so area B = 0.475. Areas A and B together make up approximately 95% of the distribution, therefore. There is thus a 95% probability of the sample mean falling within the range [108 875, 264 875], and we call this the 95% probability interval for the sample mean.

- we write this:

$$\Pr(108\,875 \leq \bar{x} \leq 264\,875) = 0.95$$

- or, in terms of the formulae we have used:

$$\Pr(\mu - 1.96\sqrt{\frac{\sigma^2}{n}} \leq \bar{x} \leq \mu + 1.96\sqrt{\frac{\sigma^2}{n}}) = 0.95$$

- The 95% probability interval and the related concept of the 95% confidence interval play important roles in statistical inference. We deliberately designed the example above to arrive at an answer of 95% for this reason.

4.6 The relationship between the Binomial and Normal distributions:

- Many statistical distributions are related to one another in some way.
- This means that many problems can be solved by a variety of different methods (using different distributions), though usually one is more convenient or more accurate than the others.
- This point may be illustrated by looking at the relationship between the Binomial and Normal distributions.
- Recall the experiment of tossing a coin repeatedly and noting the number of heads. We said earlier that this can be analysed via the Binomial distribution.
- But note that the number of heads, a random variable, is influenced by many independent random events (the individual tosses) added together. Furthermore, each toss counts equally, none dominates. These are just the conditions under which a Normal distribution arises, so it looks like there is a connection between the two distributions.

4.6 The relationship between the Binomial and Normal distributions:

- This idea is correct. Recall that if a random variable r follows a Binomial distribution then

$$r \sim B(n, P)$$

- and the mean of the distribution is nP and the variance $nP(1 - P)$.
- It turns out that as n increases, the Binomial distribution becomes approximately the same as a Normal distribution with mean nP and variance $nP(1 - P)$.
- This approximation is sufficiently accurate as long as $nP > 5$ and $n(1 - P) > 5$, so the approximation may not be very good (even for large values of n) if P is very close to zero or one.
- For the coin tossing experiment, where $P = 0.5$, 10 tosses should be sufficient.
- Note that this approximation is good enough with only 10 observations even though the underlying probability distribution is nothing like a Normal distribution.
- To demonstrate, the following problem is solved using both the Binomial and Normal distributions. Forty students take an exam in statistics which is simply graded pass/fail. If the probability, P, of any individual student passing is 60%, what is the probability of at least 30 students passing the exam?

4.6 The relationship between the Binomial and Normal distributions:

- The sample data are

$$P = 0.6$$
$$1 - P = 0.4$$
$$n = 40$$

- ***Binomial distribution method***

- To solve the problem using the Binomial distribution it is necessary to find the probability of exactly 30 students passing, plus the probability of 31 passing, plus the probability of 32 passing, etc., up to the probability of 40 passing (the fact that the events are mutually exclusive allows this).

- The probability of 30 passing is

$$\Pr(r = 30) = nCr \times P^r(1 - P)^{n-r} = 40C^{30} \times 0.6^{30} \times 0.4^{10} = 0.02$$

- This by itself is quite a tedious calculation, but $\Pr(31), \Pr(32),$ etc., still have to be calculated. Calculating these and summing them gives the result of 3.52% as the probability of at least 30 passing. (It would be a useful exercise for you to do, if only to appreciate how long it takes.)

4.6 The relationship between the Binomial and Normal distributions:

- ***Normal distribution method***
- the Binomial distribution can be approximated by a Normal distribution with mean nP and variance $nP(1-P)$. nP in this case is $24\,(40 \times 0.6)$ and $n(1-P)$ is 16, both greater than 5, so the approximation can be safely used. Thus

$$r \sim N\big(nP, nP(1-P)\big) \equiv r \sim N(24, 9.6)$$

- The usual methods are then used to find the appropriate area under the distribution.
- However, before doing so, there is one adjustment to be made (this only applies when approximating the Binomial distribution by the Normal).
- The Normal distribution is a continuous one while the Binomial is discrete.
- Thus 30 in the Binomial distribution is represented by the area under the Normal distribution between 29.5 and 30.5. 31 is represented by 30.5 to 31.5, etc.
- Thus it is the area under the Normal distribution to the right of 29.5, not 30, which must be calculated. This is known as the continuity correction.

4.6 The relationship between the Binomial and Normal distributions:

- Calculating the z score gives

$$z = \frac{29.5 - 24}{\sqrt{9.6}} = 1.78$$

- This gives an area of 3.75%, not far off the correct answer as calculated by the Binomial distribution.
- The time saved and ease of calculation would seem to be worth the slight loss in accuracy.
- Other examples can be constructed to test this method, using different values of P and n. Small values of n, or values of nP or $n(1 - P)$ less than 5, will give poor results, i.e. the Normal approximation to the Binomial will not be very good.

4.6 The relationship between the Binomial and Normal distributions:

- **Exercise 2**

- (a) A coin is tossed 20 times. What is the probability of more than 14 heads? Perform the calculation using both the Binomial and Normal distributions, and compare results.

- (b) A biased coin, for which $\Pr(H) = 0.7$ is tossed 6 times. What is the probability of more than 4 heads? Compare Binomial and Normal methods in this case. How accurate is the Normal approximation?

- (c) Repeat part (b) but for more than 5 heads.

4.7 The Poisson distribution:

- The section above showed how the Binomial distribution could be approximated by a Normal distribution under certain circumstances.
- The approximation does not work particularly well for very small values of P, when nP is less than 5.
- In these circumstances the Binomial may be approximated instead by the Poisson distribution, which is given by the formula

$$\Pr(x) = \frac{\mu^x e^{-\mu}}{x!}$$

- where μ is the mean of the distribution (similar to μ for the Normal distribution and nP for the Binomial).
- Like the Binomial, *but unlike the Normal*, the Poisson is a discrete probability distribution, so that equation above is only defined for integer values of x.
- Furthermore, it is applicable to a series of trials which are independent, as in the Binomial case.
- The use of the Poisson distribution is appropriate when the probability of ‘success’ is very small and the number of trials large.

4.7 The Poisson distribution:

- *Example:* A manufacturer gives a two-year guarantee on the TV screens it makes. From past experience it knows that 0.5% of its screens will be faulty and fail within the guarantee period.
- What is the probability that of a consignment of 500 screens (a) none will be faulty, (b) more than three are faulty?
- Solution:
- The mean of the Poisson distribution in this case is $\mu = 2.5$ $(0.5\%\ of\ 500)$. Therefore

$$\Pr(x = 0) = \frac{2.5^0 e^{-2.5}}{0!} = 0.082$$

- giving a probability of 8.2% of no failures. The answer to this problem via the Binomial method is

$$\Pr(r = 0) = 0.995^{500} = 0.0816$$

- Thus the Poisson method gives a reasonably accurate answer. The Poisson approximation to the Binomial is satisfactory if nP is less than about 7.

4.7 The Poisson distribution:

- The probability of more than three screens expiring is calculated as

$$\Pr(x > 3) = 1 - \Pr(x = 0) - \Pr(x = 1) - \Pr(x = 2) - \Pr(x = 3)$$

$$\Pr(x = 1) = \frac{2.5^1\, e^{-2.5}}{1!} = 0.205$$

$$\Pr(x = 2) = \frac{2.5^2\, e^{-2.5}}{2!} = 0.256$$

$$\Pr(x = 3) = \frac{2.5^3\, e^{-2.5}}{3!} = 0.214$$

- So

$$\Pr(x > 3) = 1 - 0.082 - 0.205 - 0.214 = 0.242$$

- Thus there is a probability of about 24% of more than three failures. The Binomial calculation is much more tedious, but gives an answer of 24.2% also.

4.7 The Poisson distribution:

- The Poisson distribution is also used in problems where events occur over time, such as goals scored in a football match or queuing type problems (e.g. arrivals at a bank cash machine).
- In these problems, there is no natural 'number' of trials but it is clear that, if we take a short interval of time, the probability of an event occurring is small. We can then consider the number of trials to be the number of time intervals.
- This is illustrated by the following example.
- A football team scores, on average, two goals every game (you can vary the example by using your own favourite team plus their scoring record!). What is the probability of the team scoring zero or one goal during a game?
- The mean of the distribution is 2, so we have, using the Poisson distribution

$$\Pr(x = 0) = \frac{2^0\, e^{-2}}{0!} = 0.135$$

$$\Pr(x = 1) = \frac{2^1\, e^{-2}}{2!} = 0.271$$

calculate the probabilities of <u>2 or more goals</u> and verify that the probabilities sum to 1.

4.7 The Poisson distribution:

- **Exercise 3**
- (a) The probability of winning a prize in a lottery is 1 in 50. If you buy 50 tickets, what is the probability that

  (i) 0 tickets win,

  (ii) 1 ticket wins,

  (iii) 2 tickets win.

  (iv) What is the probability of winning at least one prize?

- (b) On average, a person buys a lottery ticket in a supermarket every 5 minutes. What is the probability that 10 minutes will pass with no buyers?