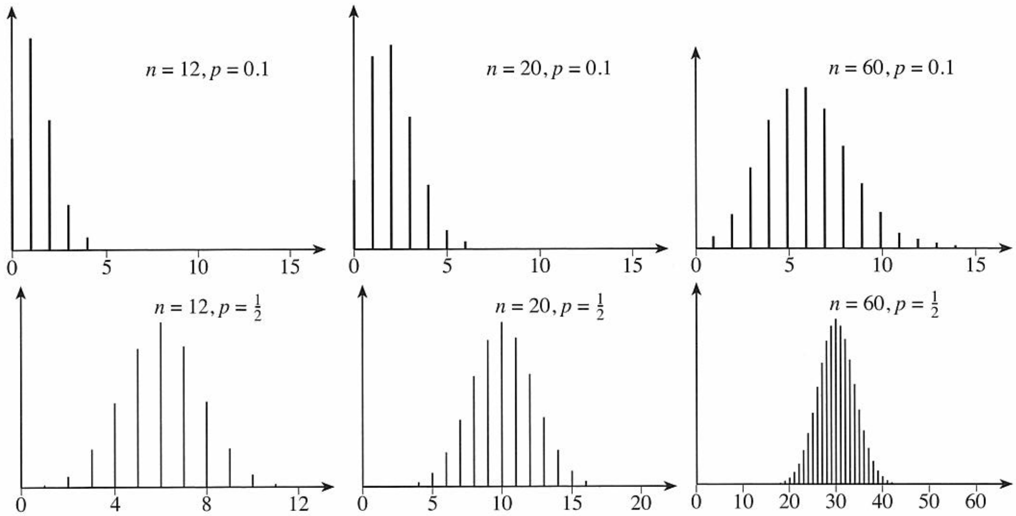


Fig. 9.18. Bar charts showing the binomial distribution for different values of n and p.

You should be able to see that if a third row of bar charts were drawn with $p = 0.9$ for the three values of n, 12, 20 and 60, then each bar chart would be the mirror image of the corresponding bar chart in the first row, about the vertical line $x = \frac{1}{2}n$. Thus, the bar chart for $n = 12$ and $p = 0.9$ is the mirror image of that for $n = 12$ and $p = 0.1$.

Summarising these observations you find that if $X \sim \mathrm{B}(n, p)$ and n is sufficiently large, then the distribution of X is approximately normal. If p is close to 0 or 1, then n must be larger than if p is close to $\frac{1}{2}$.

There is a simple condition which you can use to test whether a binomial distribution can reasonably be approximated by a normal distribution.

> If $X \sim \mathrm{B}(n, p)$, and if $np > 5$ and $nq > 5$, where $q = 1 - p$, then the distribution of X can reasonably be approximated by a normal distribution.

Notice that if $p = \frac{1}{2}$, then the conditions $np > 5$ and $nq > 5$ will be satisfied when $n > 10$. You saw that when $n = 12$ and $p = \frac{1}{2}$ the normal approximation would be reasonable.

On the other hand, if $p = 0.1$ or if $p = 0.9$, then only when $n > 50$ will *both* $np > 5$ *and* $nq > 5$ be satisfied. You saw that for a $\mathrm{B}(60, 0.1)$ distribution the normal approximation seemed satisfactory. A normal approximation would still be satisfactory for a $\mathrm{B}(60, 0.9)$ distribution because the bar chart representing this distribution is a mirror image of the one representing a $\mathrm{B}(60, 0.1)$ distribution.

If the normal distribution is a good approximation to a binomial distribution you would imagine that the normal curve would nearly pass through the tops of the bars on the bar chart. In this case it seems reasonable to suppose that the two distributions have the same mean and variance.

But you know that if $X \sim B(n,p)$, then $E(X) = np$ and $Var(X) = npq$. The approximating normal distribution should therefore also have mean np and variance npq. That is, the distribution of X is approximated by a $N(np, npq)$ distribution.

Summarising:

If $X \sim B(n,p)$, and if $np > 5$ and $nq > 5$, where $q = 1 - p$, then X can reasonably be approximated by $V \sim N(np, npq)$.

For example, if $X \sim B(60, \frac{1}{2})$, then the distribution of X can be approximated by the normal distribution with mean $\mu = 60 \times \frac{1}{2} = 30$ and variance $\sigma^2 = 60 \times \frac{1}{2} \times \frac{1}{2} = 15$. So the distribution $X \sim B(60, \frac{1}{2})$ can be approximated by the distribution $V \sim N(30,15)$.

If you have access to a computer, see whether you can draw the bar charts in Fig. 9.18 using a spreadsheet. Experiment with values of n *and* p *and check whether the bar charts look 'normally distributed' whenever the conditions* $np > 5$ *and* $nq > 5$ *are satisfied.*

Note that the normal approximation is $N(np, npq)$, *and not* $N(np, nq)$.

But how does the normal approximation work in practice? There seems immediately to be a problem since, for example,

$$P(X = 31) = \binom{60}{31} \times 0.5^{31} \times 0.5^{29} \neq 0,$$

whereas

$$P(V = 31) = 0$$

since V has a continuous distribution.

In fact, $P(V = v) = 0$ for any $v = 0, 1, 2, \ldots, 60$, so a little more work is needed in order to make the connection between the binomial distribution and the normal distribution useful.

Usually discrete distributions are represented by bar charts in which each bar is separate from the next bar and in which the height of the bar is proportional to the probability of that particular value, as in Fig. 9.19.

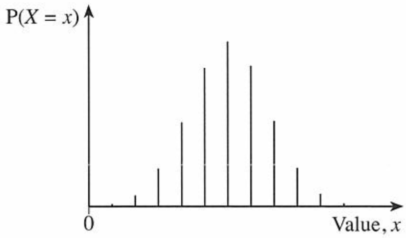


Fig. 9.19. Bar chart showing the distribution of a discrete random variable, X.

In order to compare the binomial distribution with the normal distribution it is more helpful to draw a bar chart that looks similar to a histogram. You widen each of the bars by a $\frac{1}{2}$ unit on either side so that each is still centred on an integer value but the blocks now touch. Each block has an area proportional to the probability of the integer on which it is centred. The resulting diagram is more suitable for comparison with the normal distribution as you can see in Fig. 9.20.

The possible values which a $B(n, p)$ distribution can take are $0, 1, 2, \ldots, n$, whereas the $N(np, npq)$ distribution can take all real values. In particular, this means that the normal distribution can take values greater than n or less than 0. This would seem to indicate that a normal distribution is not a satisfactory approximation to a binomial distribution. However, the area below the normal curve for x-values above n or below 0 will be very small indeed and so, although it is possible for the $N(np, npq)$ distribution to be greater than n or less than 0, it is very unlikely indeed and you do not need to worry about the fact that a normal distribution can take all real values.

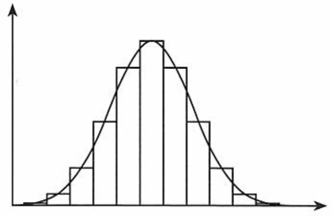


Fig. 9.20. Comparison between the binomial distribution and the normal distribution.

Fig. 9.21a and Fig. 9.21b show how you can make the relationship between the normal approximation and the binomial distribution more precise.

You can see that the probability which is represented by the shaded rectangular region in Fig. 9.21a can be approximated by the shaded region in Fig. 9.21b, which resembles a trapezium and which is the area under the normal curve between 30.5 and 31.5.

In other words, $P(X = 31) \approx P(30.5 < V < 31.5)$.

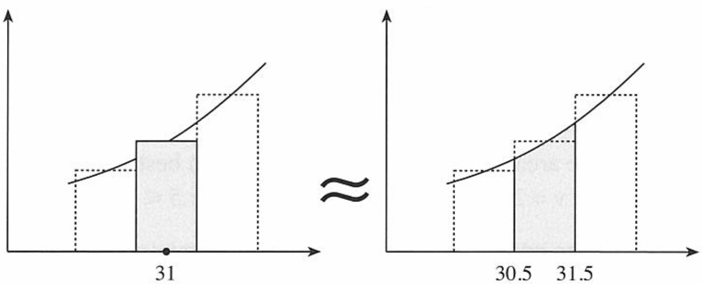


Fig. 9.21a Fig. 9.21b

The normal distribution approximates to the binomial distribution.

If you wanted to calculate $P(X \leqslant 26)$ you would calculate the total area of all the blocks up to and including the block at $x = 26$. This area is best approximated by the area under the normal curve up to $v = 26.5$. That is,

$$P(X \leqslant 26) \approx P(V \leqslant 26.5).$$

The 'extra' 0.5 which appears is necessary because you are approximating a discrete distribution by a continuous distribution. It is called a **continuity correction** and it is needed in order to make the approximation as accurate as possible.

Example 9.7.1
The random variable X has a $B(60,\frac{1}{2})$ distribution. For each of the following binomial probabilities describe the region under the approximating normal curve whose area gives the best estimate.

(a) $P(X \leqslant 12)$ (b) $P(X = 16)$ (c) $P(X < 22)$
(d) $P(X > 18)$ (e) $P(12 < X \leqslant 34)$ (f) $P(12 \leqslant X < 21)$

Since $X \sim B(60,\frac{1}{2})$, it can be approximated by the distribution of V where $V \sim N(30,15)$.

(a) To find $P(X \leqslant 12)$ you need to find the total area of the blocks for $x = 0, 1, 2, \ldots, 12$. The area under the normal curve which best approximates to this is the area for which $v \leqslant 12.5$, so $P(X \leqslant 12) \approx P(V \leqslant 12.5)$.

(b) To find $P(X = 16)$ you need to find the total area of the block for $x = 16$. The area under the normal curve which best approximates to this is the area between $v = 15.5$ and $v = 16.5$, so $P(X = 16) \approx P(15.5 \leqslant V \leqslant 16.5)$.

(c) To find $P(X < 22)$ you need to find the total area of the blocks for $x = 0, 1, 2, \ldots, 21$. The area under the normal curve which best approximates to this is the area for which $v \leqslant 21.5$, so $P(X < 22) \approx P(V \leqslant 21.5)$.

(d) To find $P(X > 18)$ you need to find the total area of the blocks for $x = 19, \ldots, 60$. The area under the normal curve which best approximates to this is the area for which $v \geqslant 18.5$, so $P(X > 18) \approx P(V \geqslant 18.5)$.

(e) To find $P(12 < X \leqslant 34)$ you need to find the total area of the blocks for $x = 13, 14, \ldots, 34$. The area under the normal curve which best approximates to this is between $v = 12.5$ and $v = 34.5$, so $P(12 < X \leqslant 34) \approx P(12.5 \leqslant V \leqslant 34.5)$.

(f) To find $P(12 \leqslant X < 21)$ you need to find the total area of the blocks for $x = 12, 13, \ldots, 20$. The area under the normal curve which best approximates to this is between $v = 11.5$ and $v = 20.5$, so $P(12 \leqslant X < 21) \approx P(11.5 \leqslant V \leqslant 20.5)$.

The best way to work with the normal approximation to the binomial distribution using the continuity correction is to draw a sketch of the situation each time it arises. Do not try to learn the results in Example 9.7.1.

Notice that there was no need in Example 9.7.1 to test the validity of the normal approximation to the binomial distribution, because it had already been verified in the discussion which preceded the example. That is not the case in the next example.

Example 9.7.2
A random variable X has a binomial distribution with parameters $n = 80$ and $p = 0.4$. Use a suitable approximation to calculate the following probabilities.

(a) $P(X \leqslant 34)$ (b) $P(X \geqslant 26)$ (c) $P(X = 33)$ (d) $P(30 < X \leqslant 40)$.

The values of np and nq are given by $np = 80 \times 0.4 = 32$ and $nq = 80 \times (1 - 0.4) = 48$. Since these are both greater than 5 the normal distribution is a good approximation to the binomial distribution, so you can approximate to $X \sim B(80,0.4)$ by $V \sim N(np,npq) = N(32,19.2)$.

Now that you are working with a normal distribution, V, standardise by letting $Z = \frac{V-32}{\sqrt{19.2}}$. Then $Z \sim \mathrm{N}(0,1)$.

(a) $\mathrm{P}(X \leqslant 34) \approx \mathrm{P}(V \leqslant 34.5) = \mathrm{P}\left(Z \leqslant \frac{34.5-32}{\sqrt{19.2}}\right) = \mathrm{P}(Z \leqslant 0.570\ldots)$

$= \Phi(0.571) = 0.7160$

$= 0.716$, correct to 3 decimal places.

(b) $\mathrm{P}(X \geqslant 26) \approx \mathrm{P}(V \geqslant 25.5) = \mathrm{P}\left(Z \geqslant \frac{25.5-32}{\sqrt{19.2}}\right) = \mathrm{P}(Z \geqslant -1.483)$

$= 1-\Phi(-1.483) = 1-(1-\Phi(1.483)) = \Phi(1.483)$

$= 0.9310 = 0.931$, correct to 3 decimal places.

(c) $\mathrm{P}(X = 33) \approx \mathrm{P}(32.5 \leqslant V \leqslant 33.5) = \mathrm{P}\left(\frac{32.5-32}{\sqrt{19.2}} \leqslant Z \leqslant \frac{33.5-32}{\sqrt{19.2}}\right)$

$= \mathrm{P}(0.114 \leqslant Z \leqslant 0.342) = \Phi(0.342) - \Phi(0.114)$

$= 0.6338 - 0.5454 = 0.0884$

$= 0.088$, correct to 3 decimal places.

(d) $\mathrm{P}(30 < X \leqslant 40) \approx \mathrm{P}(30.5 \leqslant V \leqslant 40.5) = \mathrm{P}\left(\frac{30.5-32}{\sqrt{19.2}} \leqslant Z \leqslant \frac{40.5-32}{\sqrt{19.2}}\right)$

$= \mathrm{P}(-0.342 \leqslant Z \leqslant 1.940) = \Phi(1.940) - \Phi(-0.342)$

$= \Phi(1.940) - (1-\Phi(0.342))$

$= 0.9738 - (1-0.6338) = 0.6076$

$= 0.608$, correct to 3 decimal places.

Example 9.7.3

A manufacturer of spice jars knows that 8% of the jars produced are defective. He supplies jars in cartons containing 12 jars. He supplies cartons of jars in crates of 60 cartons. In each case making clear the distribution that you are using, calculate the probability that

(a) a carton contains exactly two defective jars,

(b) a carton contains at least one defective jar,

(c) a crate contains between 39 and 44 (inclusive) cartons with at least one defective jar.

Let D be the number of defective jars in a randomly chosen carton of 12 jars. Then $D \sim \mathrm{B}(12, 0.08)$.

Since $np = 12 \times 0.08 = 0.96 < 5$ you cannot use the normal approximation. You must therefore use the binomial distribution.

(a) $\mathrm{P}(D=2) = \binom{12}{2} \times 0.08^2 \times 0.92^{10} = 0.1834\ldots$.

The probability that a carton contains exactly two defective jars is 0.183, correct to 3 decimal places.

(b) $P(D \geqslant 1) = 1 - P(D = 0) = 1 - 0.92^{12} = 0.6323\ldots$.

The probability that a carton contains at least one defective jar is 0.632, correct to 3 decimal places.

(c) Let C be the number of cartons containing at least one defective jar in a randomly chosen crate of 60 cartons.

Then C has a binomial distribution with $n = 60$. The probability of success, p, is the value found in part (b). So $p = 0.6323\ldots$.

For this binomial distribution, $np = 60 \times 0.6323\ldots = 37.94\ldots$ and $nq = 60 \times (1 - 0.6323\ldots) = 22.06\ldots$. Since they are both greater than 5, the normal approximation is valid.

The distribution of C is approximately the same as the distribution of V where $V \sim N(60 \times 0.6323\ldots, 60 \times 0.6323\ldots \times (1 - 0.6323\ldots))$. That is, the distribution of C is approximately the same as $V \sim N(37.94, 13.95)$.

So $\quad P(39 \leqslant C \leqslant 44) \approx P(38.5 \leqslant V \leqslant 44.5)$.

Let $Z = \dfrac{V - 37.94}{\sqrt{13.95}}$. Then $Z \sim N(0,1)$.

$$\begin{aligned} P(39 \leqslant C \leqslant 44) \approx P(38.5 \leqslant V \leqslant 44.5) &= P\left(\frac{38.5 - 37.94}{\sqrt{13.95}} \leqslant Z \leqslant \frac{44.5 - 37.94}{\sqrt{13.95}}\right) \\ &= P(0.150 \leqslant Z \leqslant 1.756) = \Phi(1.756) - \Phi(0.150) \\ &= 0.9604 - 0.5596 = 0.4008. \end{aligned}$$

The probability that a crate contains between 39 and 44 (inclusive) cartons with at least one defective jar is 0.401, correct to 3 decimal places.

Returning to the problem posed at the beginning of this section, you can now use what you have learned to calculate an approximation to the probability of there being more than 150 left-handed students in a school of 1000 students.

Recall that the random variable L was defined as the number of left-handed people in a randomly chosen sample of 1000 people. The distribution of L was modelled by a $B(1000, 0.16)$ distribution.

Now $np = 160$ and $nq = 840$ and both are much greater than 5, so it is valid to use the normal approximation. The distribution of L can be approximated by the random variable $V \sim N(160, 134.4)$ and therefore $P(L > 150) \approx P(V > 150.5)$.

Standardising, let $Z = \dfrac{V - 160}{\sqrt{134.4}}$. Then

$$\begin{aligned} P(V > 150.5) &= P\left(Z > \frac{150.5 - 160}{\sqrt{134.4}}\right) = P(Z > -0.819\ldots) = P(Z < 0.819\ldots) \\ &= \Phi(0.819\ldots) = 0.7935 \\ &= 0.794, \text{ correct to 3 decimal places.} \end{aligned}$$

Exercise 9D

1 State whether the following binomial distributions can or cannot reasonably be approximated by a normal distribution. Write down a brief calculation to justify your conclusion in each case.

(a) B(50,0.2) (b) B(60,0.1) (c) B(70,0.01) (d) B(30,0.7)

(e) B(40,0.9)

2 A random variable, X, has a binomial distribution with parameters $n = 40$ and $p = 0.3$. Use a suitable approximation, which you should show is valid, to calculate the following probabilities.

(a) $P(X \geqslant 18)$ (b) $P(X < 9)$ (c) $P(X = 15)$ (d) $P(11 < X < 15)$

3 The mass production of a cheap pen results in there being one defective pen in 20 on average. Use an approximation, which you should show is valid, to find, in a batch of 300 of these pens, the probability of there being

(a) 24 or more defective pens, (b) 10 or fewer defective pens.

4 A fair coin is tossed 18 times.

(a) Use the binomial distribution to find the probability of obtaining 14 heads.

(b) Use a normal approximation to find the probability of obtaining 14 heads, and to find the probability of obtaining 14 or more heads. Show that the approximation is valid.

5 In a certain country 12% of people have green eyes. If 50 people from this country are inspected, find the probability that

(a) 12 or more of them have green eyes,

(b) between 3 and 10 (inclusive) of them have green eyes.

Show that your approximation is valid.

6 Pierre attempts to dial a connection to the internet for his email each day. He is successful on his first attempt eight times out of ten. Use a normal approximation, showing first that it is valid, to find the probability that Pierre is successful on his first attempt at dialling a connection on 36 days or more over a period of 40 days.

7 (a) An unbiased dice is thrown 60 times. Find the probability that a 5 is obtained on 12 to 18 (inclusive) of these throws.

(b) In a game two unbiased dice are thrown. A winning score on each throw is a total of 5, 6, 7 or 8. Find the probability of a win on 70 or more throws out of 120 throws.

8 At an election there are two parties, X and Y. On past experience twice as many people vote for party X as for party Y.

An opinion poll researcher samples 90 voters. Find the probability that 70 or more say they will vote for party X at the next election.

If 2000 researchers each question 90 voters, how many of these researchers would be expected to record '70 or more for party X' results?

9 A manufacturer states that 'three out of four people prefer our product (Acme) to a competitor's product'. To test this claim a researcher asks 80 people about their liking for Acme. Assuming that the manufacturer is correct, find the probability that fewer than 53 prefer Acme. If 1000 researchers each question 80 people, how many of these researchers would be expected to record 'fewer than 53 prefer Acme' results?

10 Videos are packed in a box which contains 20 videos. 5% of the videos are faulty. The boxes are packed in crates which contain 50 boxes. Find the probabilities of the following events, clearly stating which distribution you are using and why.

(a) A box contains two faulty videos.

(b) A box contains at least one faulty video.

(c) A crate contains between 35 and 39 (inclusive) boxes with at least one faulty video.

Miscellaneous exercise 9

1 Given that $X \sim \text{N}(10, 2.25)$, find $\text{P}(X > 12)$. (OCR)

2 The random variable X has the distribution $X \sim \text{N}(10, 8)$. Find $\text{P}(X > 6)$. (OCR)

3 W is a normally distributed random variable with mean 0.58 and standard deviation 0.12. Find $\text{P}(W < 0.79)$. (OCR)

4 X is a random variable with the distribution $X \sim \text{N}(140, 56.25)$. Find the probability that X is greater than 128.75. (OCR)

5 The manufacturers of a new model of car state that, when travelling at 56 miles per hour, the petrol consumption has a mean value of 32.4 miles per gallon with standard deviation 1.4 miles per gallon. Assuming a normal distribution, calculate the probability that a randomly chosen car of that model will have a petrol consumption greater than 30 miles per gallon when travelling at 56 miles per hour. (OCR)

6 A normally distributed random variable, X, has mean 20.0 and variance 4.15. Find the probability that $18.0 < X < 21.0$. (OCR)

7 The lifetime of a Fotobrite light bulb is normally distributed with mean 1020 hours and standard deviation 85 hours. Find the probability that a Fotobrite bulb chosen at random has a lifetime between 1003 and 1088 hours. (OCR)

8 The area that can be painted using one litre of Luxibrite paint is normally distributed with mean 13.2 m^2 and standard deviation 0.197 m^2. The corresponding figures for one litre of Maxigloss paint are 13.4 m^2 and 0.343 m^2. It is required to paint an area of 12.9 m^2. Find which paint gives the greater probability that one litre will be sufficient, and obtain this probability. (OCR)

9 The random variable X is normally distributed with mean and standard deviation both equal to a.

Given that $\text{P}(X < 3) = 0.2$, find the value of a. (OCR)

10 The time required to complete a certain car journey has been found from experience to have mean 2 hours 20 minutes and standard deviation 15 minutes.

(a) Use a normal model to calculate the probability that, on one day chosen at random, the journey requires between 1 hour 50 minutes and 2 hours 40 minutes.

(b) It is known that delays occur rarely on this journey, but that when they do occur they are lengthy. Give a reason why this information suggests that a normal distribution might not be a good model. (OCR)

11 The weights of eggs, measured in grams, can be modelled by a $X \sim \mathrm{N}(85.0, 36)$ distribution. Eggs are classified as large, medium or small, where a large egg weighs 90.0 grams or more, and 25% of eggs are classified as small. Calculate

(a) the percentage of eggs which are classified as large,

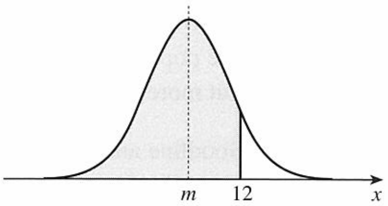
(b) the maximum weight of a small egg. (OCR)

12 The random variable X is normally distributed with standard deviation 3.2. The probability that X is less than 74 is 0.8944.

(a) Find the mean of X.

(b) Fifty independent observations of X are made. Find the expected number of observations that are less than 74. (OCR)

13 A random variable X has a $\mathrm{N}(m, 4)$ distribution. Its associated normal curve is shown in the diagram. Find the value of m such that the shaded area is 0.800, giving your answer correct to 3 significant figures. (OCR)



14 A machine cuts a very long plastic tube into short tubes. The length of the short tubes is modelled by a normal distribution with mean m cm and standard deviation 0.25 cm. The value of m can be set by adjusting the machine. Find the value of m for which the probability is 0.1 that the length of a short tube, picked at random, is less than 6.50 cm.

The machine is adjusted so that $m = 6.40$, the standard deviation remaining unchanged. Find the probability that a tube picked at random is between 6.30 and 6.60 cm long. (OCR)

15 A university classifies its degrees as Class 1, Class 2.1, Class 2.2, Class 3, Pass and Fail. Degrees are awarded on the basis of marks which may be taken as continuous and modelled by a normal distribution with mean 57.0 and standard deviation 10.0. In a particular year, the lowest mark for a Class 1 degree was 70.0, the lowest mark for a Class 2.1 degree was 60.0, and 4.5% of students failed. Calculate

(a) the percentage of students who obtained a Class 1 degree,

(b) the percentage of students who obtained a Class 2.1 degree,

(c) the lowest possible mark for a student who obtained a Pass degree. (OCR)

16 The number of hours of sunshine at a resort has been recorded for each month for many years. One year is selected at random and H is the number of hours of sunshine in August of that year. H can be modelled by a normal variable with mean 130.

(a) Given that $\text{P}(H < 179) = 0.975$, calculate the standard deviation of H.

(b) Calculate $\text{P}(100 < H < 150)$. (OCR)

17 The mass of grapes sold per day in a supermarket can be modelled by a normal distribution. It is found that, over a long period, the mean mass sold per day is 35.0 kg, and that, on average, less than 15.0 kg are sold on one day in twenty.

(a) Show that the standard deviation of the mass of grapes sold per day is 12.2 kg, correct to 3 significant figures.

(b) Calculate the probability that, on a day chosen at random, more than 53.0 kg are sold. (OCR)

18 An ordinary unbiased dice is thrown 900 times. Using a suitable approximation, find the probability of obtaining at least 160 sixes. (OCR)

19 The random variable X is normally distributed with mean μ and variance σ^2. It is given that $\text{P}(X > 81.89) = 0.010$ and $\text{P}(X < 27.77) = 0.100$. Calculate the values of μ and σ. (OCR)

20 It is given that 40% of the population support the Gamboge Party. One hundred and fifty members of the population are selected at random. Use a suitable approximation to find the probability that more than 55 out of the 150 support the Gamboge Party. (OCR)

21 Two firms, Goodline and Megadelay, produce delay lines for use in communications. The delay time for a delay line is measured in nanoseconds (ns).

(a) The delay times for the output of Goodline may be modelled by a normal distribution with mean 283 ns and standard deviation 8 ns. What is the probability that the delay time of one line selected at random from Goodline's output is between 275 and 286 ns?

(b) It is found that, in the output of Megadelay, 10% of the delay times are less than 274.6 ns and 7.5% are more than 288.2 ns. Again assuming a normal distribution, calculate the mean and standard deviation of the delay times for Megadelay. Give your answers correct to 3 significant figures. (OCR)

22 State conditions under which a binomial probability model can be well approximated by a normal model.

X is a random variable with the distribution $X \sim \text{B}(12, 0.42)$.

(a) Anne uses the binomial distribution to calculate the probability that $X < 4$ and gives 4 significant figures in her answer. What answer should she get?

(b) Ben uses a normal distribution to calculate an approximation for the probability that $X < 4$ and gives 4 significant figures in his answer. What answer should he get?

(c) Given that Ben's working is correct, calculate the percentage error in his answer. (OCR)

23 A large box contains many plastic syringes, but previous experience indicates that 10% of the syringes in the box are defective. 5 syringes are taken at random from the box. Use a binomial model to calculate, giving your answers correct to three decimal places, the probability that

(a) none of the 5 syringes is defective,

(b) at least 2 syringes out of the 5 are defective.

Discuss the validity of the binomial model in this context.

Instead of removing 5 syringes, 100 syringes are picked at random and removed. A normal distribution may be used to estimate the probability that at least 15 out of the 100 syringes are defective. Give a reason why it may be convenient to use a normal distribution to do this, and calculate the required estimate. (OCR)

24 On average my train is late on 45 journeys out of 100. Next week I shall be making 5 train journeys. Let X denote the number of times my train will be late.

(a) State one assumption which must be made for X to be modelled by a binomial distribution.

(b) Find the probability that my train will be late on all of the 5 journeys.

(c) Find the probability that my train will be late on 2 or more out of the 5 journeys.

Approximate your binomial model by a suitable normal model to estimate the probability that my train is late on 20 or more out of 50 journeys. (OCR)

25 The random variable Y has the distribution $\mathrm{N}(\mu,16)$. Given that $\mathrm{P}(Y > 57.50) = 0.1401$, find the value of μ giving your answer correct to 2 decimal places. (OCR)

26 The playing time, T minutes, of classical compact discs is modelled by a normal variable with mean 61.3 minutes. Calculate the standard deviation of T if 5% of discs have playing times greater than 78 minutes. (OCR)

27* The random variable Y is such that $Y \sim \mathrm{N}(8,25)$. Show that, correct to 3 decimal places, $\mathrm{P}(|Y-8| < 6.2) = 0.785$.

Three random observations of Y are made. Find the probability that exactly two observations will lie in the interval defined by $|Y-8| < 6.2$. (OCR)

28 It is estimated that, on average, one match in five in the Football League is drawn, and that one match in two is a home win.

(a) Twelve matches are selected at random. Calculate the probability that the number of drawn matches is

(i) exactly three,

(ii) at least four.

(b) Ninety matches are selected at random. Use a suitable approximation to calculate the probability that between 13 and 20 (inclusive) of the matches are drawn.

(c) Twenty matches are selected at random. The random variables D and H are the numbers of drawn matches and home wins, respectively, in these matches. State, with a reason, which of D and H can be better approximated by a normal variable. (OCR)

29 Squash balls, dropped onto a concrete floor from a given point, rebound to heights which can be modelled by a normal distribution with mean 0.8 m and standard deviation 0.2 m. The balls are classified by height of rebound, in order of decreasing height, into these categories: Fast, Medium, Slow, Super-Slow and Rejected.

(a) Balls which rebound to heights between 0.65 m and 0.9 m are classified as Slow. Calculate the percentage of balls classified as Slow.

(b) Given that 9% of balls are classified as Rejected, calculate the maximum height of rebound of these balls.

(c) The percentages of balls classified as Fast and as Medium are equal. Calculate the minimum height of rebound of a ball classified as Fast, giving your answer correct to 2 decimal places. (OCR)

Revision exercise

1 The table shows the length distribution of pebbles from the bed of a river.

Length, x (mm)	$0 \leqslant x < 5$	$5 \leqslant x < 10$	$10 \leqslant x < 20$	$20 \leqslant x < 50$	$50 \leqslant x < 100$
Frequency	10	8	12	25	30

(a) You are given that the frequency density for the class $0 \leqslant x < 5$ is 2. Write down the frequency densities for the other classes.

(b) Represent the data in a histogram.

(c) Calculate an estimate of the mean length of the pebbles in the sample, and of the standard deviation of the length of the pebbles in the sample.

2 The six faces of an ordinary dice are numbered 1 to 6 in the usual fashion, where the total on opposite faces adds up to 7.

(a) State, with justification, the assumption that leads to the probability of a given number appearing on the uppermost face being $\frac{1}{6}$ when the dice is rolled.

(b) Let X be the random variable representing the outcome when the dice is rolled. Calculate the variance of X.

3 A train is due to arrive at Central Station at 09.30 daily. On ten successive days the number of minutes by which the train was late were as follows.

3 0 4 –2 –3 13 8 –2 6 3

Show that the mean time of arrival was 09.33 and calculate the standard deviation.

On the assumption that the times of arrival are normally distributed with mean 09.33 and the standard deviation you have calculated, find the probability that the train will arrive

(a) at or before 09.30,

(b) more than 8 minutes late.

Comment on the assumption that the arrival times are normally distributed. (OCR)

4 A pupil conducting a coin-tossing experiment was surprised when she dropped 20 coins on to the floor and obtained only 5 heads.

(a) Calculate, using an appropriate binomial distribution, the probability that 20 fair coins dropped onto the floor at random will show exactly 5 heads.

(b) Show that the probability of obtaining either 5 heads or less, or 15 or more, when 20 coins are dropped on the floor at random is less than 5%.

(c) State the expected number of heads when 20 fair coins are dropped onto the floor.

(d) Comment on the result obtained in part (b).

5 The lengths, in cm, of 19 fern fronds are shown ordered below.

2.3, 2.6, 2.7, 2.8, 3.0, 3.1, 3.2, 3.5, 3.6, 3.8, 4.3, 4.4, 4.9, 4.9, 5.6, 5.9, 6.4, 6.8, 7.2

(a) Present these data in a simple stem-and-leaf display.

(b) Use your display to identify the median length and the interquartile range.

(c) Construct a box-and-whisker plot of these data.

6 Describe an experiment that you may have conducted to illustrate the normal distribution. Justify, with reference to features of this distribution, its use in the experiment.

The quantity of juice, in ml, that can be extracted from different sizes of oranges follows a normal distribution as given in the table.

	Mean	Variance
Small	70	49
Medium	90	σ^2

(a) What is the probability that more than 80 ml of juice can be extracted from one small orange?

(b) It is known that 5% of medium oranges produce more than 105 ml of juice. Calculate the value of σ.

(c) I buy 5 small oranges. Find the probability that at least 4 of them produce over 80 ml of juice. (OCR, adapted)

7 An experiment consists of shuffling an ordinary pack of 52 cards. Once shuffled, the top card is examined. The card can be a 'picture card', an 'even card' or an 'odd card'. None of these categories overlap. If it is a 'picture card' (there are 16 in a pack), a score of 5 is awarded. Otherwise, if it is an 'even card' (there are 20), a score of 2 is awarded. A score of zero is given for an 'odd card'. This is summarised in the table.

Card	Odd	Picture	Even
Score	0	5	2

(a) Write down the probability of a score of 5.

(b) Write down the probability distribution of the score, and calculate its mean and variance.

8 The probability that a toy balloon coming off a production line is faulty is 0.02. The balloons are put into bags containing 10 balloons.

(a) Assuming that faulty balloons occur at random, calculate the probability that a bag contains at least one faulty balloon.

The bags are packaged into boxes, each box containing 100 bags.

(b) Using a suitable approximation, estimate the probability that a box contains 90 or more bags of fault-free balloons.

9 In training, a high jumper, on average, clears a particular height once in every four attempts. The data obtained in training are to be used to model the jumper's performance in a competition.

(a) In the competition the bar is at this height. Write down an estimate of the probability that the jumper fails at the first attempt.

In this competition each competitor is allowed three attempts at each height. If a competitor fails on the first attempt at the height he is allowed a second attempt. If he fails a second time he is allowed a third attempt. After a third failure at the same height he is eliminated from the competition. Any competitor who clears a height at either his first, second or third attempt proceeds to the next height.

(b) Stating clearly any assumptions you make, calculate the probability that the jumper
   (i) succeeds at his second attempt at this height,
   (ii) proceeds to the next height.

(c) Give a reason, other than any assumption you made in part (b), why you think the probability model used above might be unsatisfactory, and say what modification you might make in an attempt to improve it.

10 Under what circumstances would you reasonably expect to be able to use the binomial distribution to model a probability distribution? When may a binomial distribution be approximated by a normal distribution?

Recent astronomical observations indicate that, of the 16 stars closest to our Sun, about half are accompanied by an orbiting planet at least the size of Jupiter.

(a) Assume that the proportion of such stars in the Galaxy is 50%. Calculate the probability that, in a group of 16 stars, exactly 8 have such a planetary system.

(b) The Pleiades are a cluster of some 500 stars. Use an appropriate approximation to determine the probability that there are between 230 and 270 (inclusive) stars in the Pleiades with accompanying planets at least the size of Jupiter. (OCR, adapted)

11 A certain type of examination consists of a number of questions all of equal difficulty. In a two-hour test the number of questions answered by a random sample of 1099 pupils is shown in the table.

No. of questions	0–4	5–9	10–14	15–19	20–24	25–29	30–34
No. of pupils	12	98	308	411	217	50	3

(a) Construct a cumulative frequency table and, on graph paper, draw a cumulative frequency graph.

(b) Use your graph to estimate the median and the interquartile range of these data.

(c) It should be possible for 3% of candidates to answer all the questions. Find how many questions the candidates should be asked to answer in the two hours.

12 The probability that a certain football club has all their first team players fit is 70%. When the club has a fully fit team it wins 90% of its home games. When the first team is not fully fit it wins 40% of its home games.

(a) Calculate the probability that it will win its next home game.

(b) Given that it did not win its last home game, find the probability that the team was fully fit.

13 Seven men and five women have been nominated to serve on a committee. The committee consists of four members who are to be chosen from the seven men and five women.

(a) In how many different ways can the committee be chosen?

(b) In how many of these ways will the committee consist of two men and two women?

(c) Assuming that each choice of four members is equally likely, find the probability that the committee will contain exactly two men. (OCR)

14 In an examination 30% of the candidates fail and 10% achieve distinction. Last year the pass mark (out of 200) was 84 and the minimum mark required for a distinction was 154. Assuming that the marks of the candidates were normally distributed, estimate the mean mark and the standard deviation. (OCR)

15 A mother has found that 20% of the children who accept invitations to her children's birthday parties do not come. For a particular party she invites twelve children but only has ten party hats. What is the probability that there is not a hat for every child who comes to the party?

The mother knows that there is a probability of 0.1 that a child who comes to a party will refuse to wear a hat. If this is taken into account, what is the probability that there will not be a hat for every child who wants one? (OCR)

Practice examination 1

Time 1 hour 15 minutes

Answer all the questions.
The use of an electronic calculator is expected, where appropriate.

1 An amateur weather forecaster has a theory about the chances of flooding affecting the region where he lives. He believes that if there are floods in one year the probability of floods again the next year is 0.7, and if there are no floods one year the probability of no floods the next year is 0.6. Last year, there were no floods in his region.

(i) Draw a tree diagram showing probabilities for floods and no floods for this year and next year, according to the weather forecaster's theory. [2]

(ii) Hence find the probability that there is flooding in exactly one of these two years. [2]

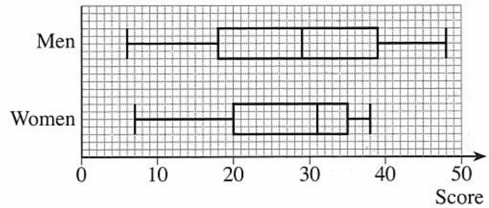
2 Two ordinary fair dice are thrown, and the random variable X denotes the larger of the two scores obtained (or the common score if the two scores are equal). The following table shows the probability distribution of X.

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

(i) Show clearly why the entry $P(X = 3) = \frac{5}{36}$ in the table is correct. [2]

(ii) Show that $E(X) = \frac{161}{36}$, and find $\text{Var}(X)$. [4]

3 The box-and-whisker plots in the diagram illustrate the scores in an aptitude test taken by people applying for a job. The scores are expressed on a scale of 0–50, and the results for men and women are shown separately.



(i) For the men taking the aptitude test, state the value of

(a) the median score, [1]

(b) the range of the scores, [1]

(c) the interquartile range of the scores. [1]

(ii) Compare briefly the scores obtained by men and women, stating one similarity and one difference. [2]

(iii) Give a reason why the scores obtained by the women would not be well modelled by a normal distribution. [1]

4 Three married couples, Mr & Mrs Lee, Mr & Mrs Martin, and Mr & Mrs Shah, stand in a line for a photograph to be taken. Find the number of different ways in which these six people can be arranged

(i) if there are no restrictions on the order in which they stand, [1]

(ii) if each man stands next to his wife, [2]

(iii) if no man stands next to another man. [3]

5 A market sells potatoes whose weights are normally distributed with mean 65 grams and standard deviation 15 grams.

(i) Find the probability that a randomly chosen potato weighs between 40 grams and 80 grams. [4]

The market sells potatoes weighing more than 80 grams separately packaged. Potatoes weighing between 80 grams and L grams are labelled as 'large' and potatoes weighing over L grams are labelled as 'extra large'.

(ii) Given that a randomly chosen potato is twice as likely to be 'large' as 'extra large', calculate the value of L. [4]

6 A survey of traffic on a busy road showed that, on average, 75% of the cars using the road carried only the driver, while 25% carried one or more passengers in addition to the driver.

(i) Twelve cars using the road are chosen at random. Find the probability that the number of these cars carrying only the driver will be

(a) exactly 9, [2]

(b) more than 9. [3]

(ii) Find the probability that more than 100 out of 120 randomly chosen cars using the road will carry only the driver. [5]

7 The values, in billions of dollars, of 100 companies registered in a certain country are summarised in the table below.

Value of company, \$$x$ billion	$1 \leqslant x < 2$	$2 \leqslant x < 3$	$3 \leqslant x < 5$	$5 \leqslant x < 10$	$10 \leqslant x < 20$
Number of companies	29	23	15	21	12

(i) Illustrate the data by means of a histogram, drawn accurately on graph paper. [4]

(ii) Calculate an estimate of the mean value of these companies, and explain briefly why your answer is only an estimate of the true mean value. [4]

(iii) The median value of x for these companies is known to be 2.92. State what feature of the data accounts for the mean being considerably greater than the median. [1]

(iv) Explain briefly why the mean might not be considered as a very good measure of the 'average' value of the companies. [1]

Practice examination 2

Time 1 hour 15 minutes

Answer all the questions.
The use of an electronic calculator is expected, where appropriate.

1 The weights, X kg, of 10-year-old boys in a certain country may be assumed to be normally distributed. The proportion of boys weighing less than 25 kg and the proportion of boys weighing more than 35 kg are each 35%.

(i) Write down the mean of X. [1]

(ii) Calculate the standard deviation of X. [3]

2 In a class of 20 pupils, seven are left-handed and 13 are right-handed. Five pupils are selected at random from the class; the order in which they are chosen is not important.

(i) Find the number of possible selections in which two of the five are left-handed and three are right-handed. [3]

(ii) Find the probability that the sample of five will include exactly two who are left-handed. [2]

3 The random variable X takes the values 0, 1, 2, 3 only, and its probability distribution is shown in the following table.

x	0	1	2	3
$\mathrm{P}(X = x)$	a	b	0.2	0.05

(i) Show that $a + b = 0.75$. [1]

(ii) Given that $\mathrm{E}(X) = 1$, find the value of b and deduce the value of a. [3]

(iii) Does X have a binomial distribution? Give reasons for your answer. [2]

4 Ali can travel to work either by bus or in his car. The probability that Ali is late for work when he goes by bus is 0.15, and the probability that he is late when he uses his car is 0.1. Ali uses his car for 70% of his journeys to work.

(i) Find the probability that Ali will be late for work on a randomly chosen day. [3]

(ii) Find the conditional probability that Ali travels by bus, given that he is late for work. [3]

5 The organisers of a TV game show think that the probability of any contestant winning a prize will be 0.7, and that the success or failure of any contestant will be independent of the success or failure of other contestants. Six contestants take part on each episode of the show.

(i) Find the probability that, in one episode of the show, the number of successful contestants will be at most 2. [4]

(ii) The show runs for 30 episodes altogether. Find the probability that the total number of successful contestants will be more than 120. [5]

6 A chicken farmer fed 25 new-born chicks with a new variety of corn. The stem-and-leaf diagram below shows the weight gains of the chicks after three weeks.

36	9
37	6
38	4 5 6
39	3 3 7 9 9
40	2 3 7 8
41	0 2 6 6
42	3 5 7
43	2 4
44	5
45	1

Key: 39|3 means 393 grams

(i) Find the median weight gain, and find also the interquartile range. [3]

The data may be summarised by $\sum(x-400)=192$ and $\sum(x-400)^2=11\,894$, where x grams is the weight gain of a chick.

(ii) Calculate the mean and standard deviation of the weight gains of the 25 chicks, giving each answer to the nearest gram. [4]

(iii) Chicks fed on the standard variety of corn had weight gains after three weeks with mean 392 grams and standard deviation 12 grams. State briefly how the new variety of corn compares to the standard variety. [2]

7 Each car owner in a sample of 100 car owners was asked the age of his or her present car. The results are shown in the table below.

Age of car, t years	$0 \leqslant t < 2$	$2 \leqslant t < 4$	$4 \leqslant t < 6$	$6 \leqslant t < 10$	$10 \leqslant t < 15$	$t \geqslant 15$
Number of cars	25	32	20	12	7	4

(i) Making a suitable assumption about the ages of the oldest cars, draw a cumulative frequency graph on graph paper to illustrate the data. [4]

(ii) Hence find estimates for

(a) the proportion of cars in the sample that are more than 5 years old, [2]

(b) the age which is exceeded by the oldest 15% of cars. [2]

(iii) Would the assumption made in part (i) have any effect on an estimate of

(a) the median age of the cars,

(b) the mean age of the cars?

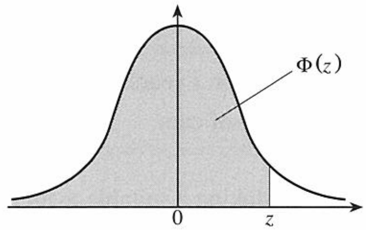
Give reasons for your answers. [3]

The Normal Distribution Function

If Z has a normal distribution with mean 0 and variance 1 then, for each value of z, the table gives the value of $\Phi(z)$, where

$$\Phi(z) = \mathrm{P}(Z \leqslant z).$$

For negative values of z use $\Phi(-z) = 1 - \Phi(z)$.



z	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
															ADD				
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	4	8	12	16	20	24	28	32	36
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	4	8	12	16	20	24	28	32	36
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	4	8	12	15	19	23	27	31	35
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	4	7	11	14	18	22	25	29	32
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	4	7	11	14	18	22	25	29	32
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	3	7	10	14	17	20	24	27	31
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	3	7	10	13	16	19	23	26	29
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	3	6	9	12	15	18	21	24	27
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	3	5	8	11	14	16	19	22	25
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	3	5	8	10	13	15	18	20	23
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	2	5	7	9	12	14	16	19	21
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	2	4	6	8	10	12	14	16	18
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	2	4	6	7	9	11	13	15	17
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	2	3	5	6	8	10	11	13	14
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	1	3	4	6	7	8	10	11	13
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	1	2	4	5	6	7	8	10	11
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	1	2	3	4	5	6	7	8	9
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	1	2	3	4	4	5	6	7	8
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	1	1	2	3	4	4	5	6	6
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	1	1	2	2	3	4	4	5	5
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	0	1	1	2	2	3	3	4	4
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	0	1	1	2	2	2	3	3	4
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	0	1	1	1	2	2	2	3	3
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	0	1	1	1	1	2	2	2	2
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	0	0	1	1	1	1	1	2	2
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	0	0	0	1	1	1	1	1	1
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	0	0	0	0	1	1	1	1	1
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	0	0	0	0	0	1	1	1	1
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	0	0	0	0	0	0	0	1	1
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	0	0	0	0	0	0	0	0	0

Critical values for the normal distribution

If Z has a normal distribution with mean 0 and variance 1 then, for each value of p, the table gives the value of z such that $\mathrm{P}(Z \leqslant z) = p$.

p	0.75	0.90	0.95	0.975	0.99	0.995	0.9975	0.999	0.9995
z	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

Answers

1 Representation of data

Exercise 1A (page 8)

1 (a) 4.3, 5.0, 5.3, 5.4, 5.7, 5.9, 6.1, 6.2, 6.3, 6.4, 7.1, 7.6, 7.6, 9.2, 9.3

(b) (i) Quantitative (ii) Continuous

2 (a)

0	4 6	(2)
1	2 5 8	(3)
2	1 5 5 5 7 8 9	(7)
3	0 2 4 6 7	(5)
4	1 3	(2)
5	2	(1)

Key: 2|7 means 27 k.p.h.

(b)

0	3 4 7 8 9 9 9	(7)
1	0 1 2 6 8	(5)
2	1 1 3 7	(4)
3		(0)
4	2	(1)

Key: 2|3 means 2.3 hours

3

3	0 4 6	(3)
4	8 8	(2)
5	0 0 1 2 3 4 4 6 6 7 8 8 9	(13)
6	0 2 3 4 7 7 8 8 9	(9)
7	0 1 4 4 4 5 5 6 7 9	(10)
8	1 6	(2)
9	1 3 9	(3)

Key: 7|9 means 79 years

4

13		(0)
13	7	(1)
14	1 2 3	(3)
14	5	(1)
15	1 1 2 3 4	(5)
15	5 6 6 7 9 9	(6)
16	0 0 0 1 1 1 2 3 4	(9)
16	5 6 7 7 8 8 9 9	(8)
17	0 1 1 1 1 2 3 4	(8)
17	5 6 7 7 8 9	(6)
18	0 0 1 1 1 2 2 3 4	(9)
18	6 6 8 9	(4)

Key: 17|9 means 179

5

2996	2 5	(2)
2997	2 4 5 6 9	(5)
2998	0 1 3 4 5 7 8 8	(8)
2999	0 1 3 3 4 6 7 8	(8)
3000	0 7	(2)

Key: 2997|4 means 299.74 thousand km s^{-1}

6 (a)

132	2 9	(2)
133	2 6 8 9	(4)
134	1 1 2 2 2 4 5 6 7 7 7 8 9	(13)
135	0 1 1 3 3 3 4 4 6	(9)
136	2	(1)
137	0	(1)

Key: 134|7 means 1.347 kg

(b) There would be only one stem, which would have all the leaves.

Exercise 1B (page 15)

B means 'class boundaries', H means 'bar heights', F means 'frequency', FD means 'frequency density'.

1 B: 45 60 75 90 105 120 150
H: 0.8 2.1 3.7 4.8 1.3 0.3

2 B: 4.5 9.5 14.5 19.5 24.5 29.5 34.5 44.5
H: 0.4 1.0 1.6 2.8 3.4 2.2 0.3

3 B: 129.5 139.5 149.5 159.5 169.5 179.5 189.5
H: 1 4 11 17 14 13

4 (a) 0, 2.5 and 2.5, 5.5

(b) B: 0 2.5 5.5 8.5 11.5 15.5
FD: 6.8 2 1.33 0.67 0.25

5 B: –0.5 9.5 19.5 29.5 34.5 39.5 49.5 59.5
H: 0.6 2.1 5.1 7.2 9.6 8.2 3.1

6 Assuming data correct to 1 d.p.,
B: 8.95 9.95 10.95 11.95 12.95 13.95 14.95 15.95 16.95
H: 1 4 7 6 9 8 7 3

7 (a) B: 2.875 3.875 4.875 5.875 6.875 7.875 8.875
H: 3 4 6 7 10 4

(b) 2.875, 3.875

(c) B: 2.875 3.875 4.875 5.875 6.875 7.875 8.875
H: 3 4 6 7 10 4

8 (a), (b) B: 16 20 30 40 50 60 80
FD: 3 4 4.4 4.7 3.2 1.25

Exercise 1C (page 18)

1 Plot at (45,0), (60,12), (75,44), (90,100), (105,172), (120,192), (150,200).

(a) 26% (b) About 77 k.p.h

2 Plot at (–0.5,0), (9.5,6), (19.5,27), (29.5,78), (34.5,114), (39.5,162), (49.5,244), (59.5,275).

(a) 28% or 29% (b) 24

3 (a) Plot at (0,0), (16,14.3), (40,47.4), (65,82.7), (80,94.6), (110,100).

(b) 10(.4) million

4 Assuming x is correct to 2 d.p., plot at $(0,0)$, $(2,15)$, $(3,42)$, $(4,106)$, $(5,178)$, $(6,264)$, $(7,334)$, $(8,350)$, $(10,360)$; about 58 poor days and 14 good days.

5 Assuming x is correct to the nearest km, plot at $(0,0)$, $(4.5,12)$, $(9.5,41)$, $(14.5,104)$, $(19.5,117)$, $(24.5,129)$, $(34.5,132)$.
(a) 8 km (b) 14 km

6 Plot at $(7,10)$, $(7.05,73)$, $(7.10,150)$, $(7.15,215)$, $(7.20,245)$, $(7.30,250)$; between 7.012 cm and 7.175 cm.

7 Plot at $(2.95,7)$, $(3.95,62)$, $(4.95,134)$, $(5.95,144)$, $(6.95,148)$, $(8.95,150)$; 9%; about 3.07.

Miscellaneous exercise 1 (page 20)

1

0	6 0 8 0	(4)
1	8 9 2 7 4 1 1 6	(8)
2	7 8 5 6 0 1 9	(7)
3	8 1 7 3 6 4	(6)
4	5 3 2 2	(4)
5	7 5	(2)
6	6 3 2	(3)
7	2 5	(2)
8	5 4 6 2	(4)

Key: 4 | 3 means 43

The diagram indicates how the scores are distributed. But it does not indicate the order in which the scores occurred.

2 B: 0 30 60 120 180 240 300 360 480
FD: 0.07 0.1 0.13 0.27 0.7 0.42 0.3 0.1
About 286 s, obtained from the cumulative frequency graph or by proportion.

3 Plot at $(100,0)$, $(110,2)$, $(120,12)$, $(130,34)$, $(140,63)$, $(150,85)$, $(160,97)$, $(180,100)$; 123 cm.

4 (a) B: 3.95 5.95 7.95 9.95 11.95 13.95 15.95
F: 3 3 4 11 8 1
(b) Plot at $(3.95,0)$, $(5.95,3)$, $(7.95,6)$ $(9.95,10)$, $(11.95,21)$, $(13.95,29)$, $(15.95,30)$; 11.4; 3.3%.

5 Groups: 1.0–2.4, 2.5–3.9, 4.0–5.4, 5.5–6.9, 7.0–8.4, 8.5–9.9, 10.0–11.4, 11.5–12.9
F: 13 16 18 15 6 8 2 2
B: 0.95 2.45 3.95 5.45 6.95 8.45 9.95 11.45 12.95
H: 13 16 18 15 6 8 2 2
For example: most times are between 1 and 7 seconds.

6 (a) Plot at $(10,16)$, $(20,47)$, $(30,549)$, $(40,1191)$, $(50,2066)$, $(60,2349)$, $(80,2394)$, $(100,2406)$, $(140,2410)$.
(b) About 74%
(c) End boundaries unknown. Use (say) 4–10, 100–140.

7 (a) Using 6 equal classes,
B: 12.75 14.25 15.75 17.25 18.75 20.25 21.75
H: 1 2 2 2 7 11
Although data appear to be correct to 2 d.p., having boundaries 12.745–14.245 would be awkward.
(b) Plot at $(12.75,0)$, $(14.25,1)$, $(15.75,3)$, $(17.25,5)$, $(18.75,7)$, $(20.25,14)$, $(21.75,25)$.
(c)

12	76	(1)
13		(0)
14	82	(1)
15	61	(1)
16	53, 97	(2)
17	30, 71	(2)
18		(0)
19	12, 35, 41, 61, 72	(5)
20	02, 21, 27, 34, 40, 52, 57, 69	(8)
21	04, 13, 25, 38, 43	(5)

Key: 16 | 53 means 16.53

8 (a) F: 17, 11, 10, 9, 3, 4, 2, 4
(b) B: –0.5 9.5 19.5 29.5 39.5 49.5 59.5 69.5 99.5
FD: 1.7, 1.1, 1.0, 0.9, 0.3, 0.4, 0.2, 0.13
(c) Plot at $(-0.5,0)$, $(9.5,17)$, $(19.5,28)$, $(29.5,38)$, $(39.5,47)$, $(49.5,50)$, $(59.5,54)$, $(69.5,56)$, $(99.5,60)$; 42.
(d) It assumes the data are evenly spread over the class 30–39. There are two each of 31, 33 and 39, and one each of 32, 36, 37, so the assumption is not well founded.

9 (a) Street 1: Plot at $(61,0)$, $(65,4)$, $(67,15)$, $(69,33)$, $(71,56)$, $(73,72)$, $(75,81)$, $(77,86)$, $(79,90)$, $(83,92)$.
Street 2: Plot at $(61,0)$, $(65,2)$, $(67,5)$, $(69,12)$, $(71,24)$, $(73,51)$, $(75,67)$, $(77,77)$, $(79,85)$, $(83,92)$.
(b) 69.8 dB on Street 1, 72.3 dB on Street 2.
(c) Street 1 appears less noisy, in general, than Street 2. For example, there are 56 readings under 71 dB for Street 1, but 24 for Street 2.

10

0	2 2 3 4 5 5 6 8	(8)
1	0 1 2 2 2 3 4 6 6 6 9	(11)
2	0 2 3 4 4 4 5 5 9	(9)
3	4 5 9	(3)
4	0 1 4 8	(4)
5	0 6 8	(3)
6	1 6 7 7	(4)
7	2 6	(2)
8	2 5	(2)
9		(0)
10	4	(1)
11	8 8 9	(3)

Key: 4|8 means 48

Assuming data correct to the nearest second,
B: 0 19.5 39.5 59.5 79.5 99.5 119.5
FD: 0.97 0.6 0.35 0.3 0.1 0.2

11 (a) 4 cm (b) 7 (c) 3.0

2 Measures of location

Exercise 2A (page 26)

1 5.4 kg; 5.7 kg

2 40

3 (a) 27.5 k.p.h (b) 1.1 hours

4 13.5

5 \$500 approximately

6 4.7 s; 5.0 s, data not evenly spread over class 4.0–5.4.

7 34

8 (a) 78 kg (b) 63 kg
On average men have greater mass than women.

Exercise 2B (page 30)

1 10.5

2 181 cm. Students appear taller, on average than the population. This can be explained by the large values 192, 192, 194 and 196.

3 (a) 11.3 (b) 105.5

4 0.319

5 3.59

6 24.3

7 Assuming the end boundary is 150, 88.8 k.p.h.

8 (a) 0–2.5, 2.5–5.5, 5.5–8.5, 8.5–11.5, 11.5–15.5 (b) 3.56 minutes

9 \$12.89; \$12.39

10 503.46 ml

Exercise 2C (page 35)

1 (a) 0 (b) No mode (c) 2–3 (d) Brown

2 (a) Mode (b) Mean (c) Median

3 It could be true for mean or mode, not the median.

4 (a) Roughly symmetrical
(b) Skewed
(c) Skewed
(d) Roughly symmetrical

5 (a) Mean 4.875, median 5, mode 6. The data set is too small for the mode to give a reliable estimate of location. The median gives a better idea of a 'typical' mark.
(b) It has a 'tail' of low values.

Miscellaneous exercise 2 (page 36)

1 (a) There would be no entries for stems 12, 13, 14 and 15.

(b)

4	0 0 0 1 2 2 2 6 6 8 8 9 9 9	(14)
5	0 2 2 4 7 9	(6)
6	4 5 6	(3)
7	3 5 6	(3)
8	5	(1)
9	2 4	(2)
10		(0)
11	2	(1)

HI 1.66; Key: 5|3 means \$0.53

(c) \$0.52; \$0.617; modes: \$0.40, \$0.42, \$0.49
(d) Median, since it is not affected by extreme values, or mean, since it involves all values.
(e) \$0.236

2 (a) 2.56; exact (b) Both 2

3 (a) Both 22.0–23.9
(b) Not supported since modal classes the same. Either mean_1 (23.21 °C) is greater than mean_2 (22.43 °C) or median_1 ($\approx$ 23 °C) is greater than median_2 ($\approx$ 22 °C).

4 (a) Histogram
(b) Data inaccurate, grouped data, may not be evenly spread over classes.
(c) About 69 s
(d) No change in median, mean increases.

5 (a) B: −0.5 29.5 39.5 49.5 59.5 69.5 79.5 89.5 100.5
FD: 0.4, 0.7, 1.3, 2.5, 4.6, 7.8, 10.5, 2.9
(b) 72.5 (c) About 77
The marks are higher for mechanics, indicating better performance.

6 Change by +0.06 or – 0.06 depending on the order of the frequencies.

7 (a) About 20.4 hours
(b) Individual values unknown, data inaccurate, data may not be evenly distributed over the classes.
(c) 21.0 hours
(d) It has a 'tail' of low values.

8 $\bar{x} = 3.6$, me $= 4$, mo $= 3$ and mean is between mode and median.

9 (a) $\bar{m} = 453.9$, $\bar{t} = 462.9$, $\bar{d} = 9.0$; yes
(b) me $m = 294.5$, me $t = 266.5$, me $d = 5$; no

10 (a) There should be no spaces between bars; areas are not proportional to frequencies; incorrect scale on vertical axis.
(b) B: 4.5 9.5 12.5 15.5 18.5 28.5
FD: 2.8, 6.0, 5.0, 1.3, 0.8
(c) 12.9 m

3 Measures of spread

Exercise 3A (page 47)

1 (a) 17, 9.5 (b) 9.1, 2.8

2 2.2

3 3

4 \$15 420, \$23 520

5 Street 1: 70.1 dB, 4.8 dB
Street 2: 72.6 dB, 4.6 dB
Street 2 is usually noisier, with greater variation.

6 Monday: $Q_1 = 105$, $Q_2 = 170$, $Q_3 = 258$
Wednesday: $Q_1 = 240$, $Q_2 = 305$, $Q_3 = 377$
Wednesday has greater audiences in general, with less variation.

7 Fat content 0: $Q_1 = 41$, $Q_2 = 53$, $Q_3 = 61$; interquartile range (IQR) $= 20$; range $= 65$
Fat content 1: $Q_1 = 30$, $Q_2 = 37.5$, $Q_3 = 49$; IQR $= 19$; range $= 46$
Fat content 0 has generally higher rating; has greater spread at extremes.

8 (a) Negative skew (b) Positive skew
(c) Roughly symmetrical

9 Box-and-whisker plots are preferred since they give visual comparison of the shapes of distributions, the quartiles, IQRs and ranges. Histograms will indicate the general shape of the distributions and will give only a rough idea of quartiles and so on. However, means and standard deviations can be estimated from a histogram but not from a boxplot.

10 (a) \$38.73, \$43.23, \$49.24, \$54.15, \$58.42
(c) Slight negative skew

11 (a) and (b); fences at 14 and 126

12 (a) Fences at 24 and 104
(b) No outliers since all values lie inside fences.
(c) Roughly symmetrical

Exercise 3B (page 53)

1 Mean: (a) 4 (b) 5
SD: (a) 2 (b) 4.899

2 (a) 3.489 (b) 4.278

3 50.728 g, 10.076 g^2

4 ± 1.2

5 149.15 cm, 5.33 cm

6 (a) Anwar is better; his mean of 51.5 is greater than Qasim's 47.42 (Anwar scores more runs than Qasim).
(b) Qasim is more consistent since his standard deviation of 27.16 is less than Anwar's of 34.13.

7 $\bar{f} = 62.3$ kg, SD (female) is 7.34 kg,
$\bar{m} = 75.6$ kg, SD (male) is 8.45 kg
Both distributions have negative skew. Females are lighter than males by about 13 kg on average and less variable than males.

Exercise 3C (page 57)

1 0.740, 1.13

3 797.4 min^2

4 250.77(5) g, 3.51 g
Increase the number of classes; weigh more accurately; use more packets.

5 0.7109

6 −17.2, 247.36

7 135.7 cm, 176.51 cm^2

8 Mid-class values are 18.5, 23.5, 28.5, 33.5, 38.5, 46, 56, 66. Mean 37.49 years, SD 11.86 years. In the second company the general age is lower and with smaller spread.

Miscellaneous exercise 3 (page 61)

1 23.16 cm, 1.32 cm; 22.89 cm, 1.13 cm
House sparrows have smaller variability; little difference in means.

2 (a) 18.69 m, 36.20 m^2
(b) 18.92 m, 36.20 m^2

3 (a) Set B (b) Set B (c) 39.5 g, 125 g^2
(d) Individual data values not given.

4 (b) 0.060 cm
(c) 1.34 approx., close to 1.3, so the distribution is roughly symmetrical.

5 (a) The standard deviation is zero, which implies that all data values are equal.
(b) Ali caught 12.84 kg, Les 12.16 kg and Sam 2 kg, so Ali won. (c) 3.12 kg

6 (a) 1.854 cm; 1.810 cm, 1.886 cm
(b) Negative skew (c) 1.850 cm, 0.069 cm
(d) −0.18, indicating negative skew

7 (a) 8.5 minutes, 9.25 minutes
(c) 10 minutes
(d) (i) and (iii) not true, (ii) and (iv) true

8 (a) 26.9 years, 13.0 years (b) 22.4 years
Median preferred since distribution skewed, more information given by median.

4 Probability

Exercise 4A (page 71)

1 (a) $\frac{1}{2}$ (b) $\frac{2}{3}$ (c) $\frac{1}{2}$ (d) $\frac{1}{2}$ (e) $\frac{1}{6}$
(f) $\frac{5}{6}$ (g) $\frac{2}{3}$

2 (a) $\frac{1}{2}$ (b) $\frac{3}{13}$ (c) $\frac{5}{13}$ (d) $\frac{5}{26}$ (e) $\frac{9}{13}$

3 (a) $\frac{1}{6}$ (b) $\frac{5}{12}$ (c) $\frac{5}{12}$ (d) $\frac{25}{36}$
(e) $\frac{11}{36}$ (f) $\frac{5}{18}$ (g) $\frac{1}{6}$ (h) $\frac{1}{2}$

4 (a) $\frac{1}{5}$ (b) $\frac{2}{5}$ (c) $\frac{1}{3}$ (d) $\frac{1}{2}$

5 (a) $\frac{9}{25}$ (b) $\frac{4}{25}$ (c) $\frac{12}{25}$ (d) $\frac{21}{25}$ (e) $\frac{3}{5}$

Exercise 4B (page 78)

1 (a) $\frac{1}{3}$ (b) $\frac{2}{15}$ (c) $\frac{8}{15}$ (d) $\frac{13}{15}$ (e) $\frac{3}{5}$
No

2 (a) $\frac{11}{221}$ (b) $\frac{10}{17}$ (c) $\frac{7}{17}$ (d) $\frac{77}{102}$

3 (a) 0.27 (b) 0.35 (c) 0.3375

4 (a) $\frac{8}{15}$ (b) $\frac{7}{15}$ (c) $\frac{3}{5}$ (d) $\frac{2}{5}$ (e) $\frac{9}{16}$
(f) Yes (g) No

5 (a) 0.24 (b) 0.42 (c) 0.706

6 (a) 0.12 (b) 0.44 (c) 0.048 (d) 0.34
(e) 0.03 (f) 0.07 (g) 0.32

7 (a) $\frac{1}{16}$ (b) $\frac{15}{16}$ (c) $\frac{671}{1296}$

8 0.491

9 0.5073 (Note that this is bigger than 50%.)

10 0.75, 0.8

12 0.0317. Very small; the test has to be much more reliable for it to give any reliable evidence about a rare disease.

13 0.000 998. The further piece of evidence is not enough to make the prisoner's guilt at all likely; it would have to be much more certain than it is.

Miscellaneous exercise 4 (page 80)

1 (b) $\frac{2}{3}$

2 (a) 0.58 (b) 0.6

3 (a) $\frac{1}{2}$ (b) $\frac{5}{11}$

4 (a) 20% (b) 10%

5 (b) (i) $\frac{23}{189}$ (ii) $\frac{166}{189}$

6 (a) $\frac{3}{20}$ (b) $\frac{9}{35}$ (c) $\frac{7}{12}$

7 (b) $\frac{9}{26}$

8 (a) (i) $\frac{1}{5}$ (ii) $\frac{5}{13}$ (iii) $\frac{17}{25}$ (iv) $\frac{1}{2}$
(b) $\frac{21}{25}$

9 (a) $\frac{1}{2}$ (b) (i) $5p$ (ii) $4p$ (c) $\frac{1}{40}$

10 $\frac{1}{4}$
(a) 0.0577 (b) 0.1057 (c) 0.6676

11 (a) $\frac{3}{253}$ (b) $\frac{43}{138}$ (c) $\frac{11}{138}$ (d) $\frac{11}{69}$

12 (a) 0.32 (b) 0.56; 8

13 (a) (i) $\frac{3}{8}$ (ii) $\frac{4}{15}$
(b) (i) $\frac{27}{125}$ (ii) $\frac{8}{125}$ (iii) $\frac{38}{125}$

14 (a) $\frac{1}{8}$ (b) $\frac{3}{8}$ (c) $\frac{8}{9}$

15 0.017; $\frac{2}{3}$

16 (a) 0.030 (b) 0.146 (c) 0 (d) 0.712

17 (a) $\dfrac{4n-3}{n(n+1)}$ (b) $\dfrac{9}{4n-3}$

5 Permutations and combinations

Exercise 5A (page 89)

1 5040

2 24

3 120

4 720

6 119, $\frac{1}{120}$

7 50 400

8 (a) 720 (b) 48

9 64, $\frac{21}{32}$

Exercise 5B (page 93)

1 22 100

2 215 760

3 (a) 56 (b) 48

4 70, 63

5 0.25

6 0.0128

7 (a) 0.222 (b) 0.070 (c) 0.112 (d) 0.180

Exercise 5C (page 96)

1 (a) 3 632 428 800 (b) 259 459 200
(c) 39 916 800 (d) 457 228 800

2 (a) 1024 (b) 210 (c) 0.205

3 (a) 5.346×10^{13} (b) 3.097×10^{12}
(c) 0.0579

4 (a) 13! (b) 43 545 600
(c) 609 638 400

5 (a) 1260 (b) 540 (c) 300 (d) 120

6 (a) 39 916 800 (b) 20 736
(c) 1 814 400

7 (a) 12 (b) 115

Miscellaneous exercise 5 (page 97)

1 3 628 800, 45

2 (a) 40 320 (b) 1152

3 83 160

4 360

5 (a) 1440 (b) 2880

6 (a) 240 (b) 480

7 210

8 151 200

9 210, $\frac{2}{7}$

10 (a) 0.112 (b) 0.368

11 (a) 360 (b) 60

12 432

13 (a) 24 (b) 120; 2880

14 (a) 70 (b) $\frac{1}{35}$

15 (a) (1,1,8), (1,2,7), (1,3,6), (1,4,5), (2,2,6), (2,3,5) (2,4,4), (3,3,4) (b) $\frac{1}{4}$

16 1260

17 (a) 60 (b) 5

18 0.0109; the player might conclude that the deals were not random.

19 One answer would be $4!\times23!=24!$.

6 Probability distributions

Exercise 6A (page 104)

1

x	0	1	2	3	4
$P(X=x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

2

d	0	1	2	3	4	5
$P(D=d)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

3

x	1	2	3	6	10
$P(X=x)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

4

h	1	2	3	4	5	6
$P(H=h)$	$\frac{23}{36}$	$\frac{7}{36}$	$\frac{3}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

5

m	1	2	3	4	6
$P(M=m)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{2}{16}$
	8	9	12	16	
	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	

6

Number	0	1	2
Probability	$\frac{5}{12}$	$\frac{1}{2}$	$\frac{1}{12}$

7

c	1	2	3	4
$P(C=c)$	$\frac{1}{13}$	$\frac{16}{221}$	$\frac{376}{5525}$	$\frac{4324}{5525}$

8

Score	3	4	5	6	7	8
Probability	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{15}{216}$	$\frac{21}{216}$
	9	10	11	12	13	14
	$\frac{25}{216}$	$\frac{27}{216}$	$\frac{27}{216}$	$\frac{25}{216}$	$\frac{21}{216}$	$\frac{15}{216}$
	15	16	17	18		
	$\frac{10}{216}$	$\frac{6}{216}$	$\frac{3}{216}$	$\frac{1}{216}$		

Exercise 6B (page 106)

1 $\frac{1}{20}$

2 0.3

3 0.15

4 $\frac{1}{8}$

5

x	1	2	3	4	5	6
$P(X=x)$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{12}$

6 $\frac{1}{21}$

7 $\frac{20}{49}$

8 0.2

Exercise 6C (page 107)

1 (a) 130 (b) 40 (c) 120
(d) 160 (e) 360

2 (a) 105 (b) 105 (c) 245

3 12, 31, 34, 18, 5, 0 (0 is better than 1 because it makes the total 100)

4 0.468, 103

Miscellaneous exercise 6 (page 108)

1

Number	0	1	2	3
Probability	$\frac{248}{1105}$	$\frac{496}{1105}$	$\frac{304}{1105}$	$\frac{57}{1105}$

2 (a)

x	0	1	2	3	4	6
$P(X=x)$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{36}$

(b) 120

3 (a) $\frac{1}{18}$ (b) $\frac{17}{36}$ (c) $\frac{11}{17}$

4 (b) 8.3

5 $\frac{1}{8}$

7 The binomial distribution

Exercise 7A (page 114)

1 (a) 0.0819 (b) 0.0154 (c) 0.0001

2 (a) 0.2561 (b) 0.2048 (c) 0.0005

3 (a) 0.2119 (b) 0.4728 (c) 0.0498

4 (a) 0.0017 (b) 5

5 (a) 0.2461 (b) 0.4102 (c) 0.0195
(d) 0.9102

6 (a) 0.0781 (b) 0.0176

7 It avoids the calculation of $n!$ or $\binom{n}{r}$ for large values of n.

8 Equality occurs when $(n+1)p$ is an integer.

Exercise 7B (page 117)

1 (a) 0.650
(b) The students are not chosen independently.

2 0.055; no (the outcomes are still green and not-green).

3 0.114; 0.223; breakages are not independent of each other (if one egg in a box is broken, it is more likely that others will be).

4 0.065; for example: P(hurricane) is constant for each month; hurricanes occur independently of each other; P(hurricane) may not be constant for each month but dependent on the time of year.

5 0.204. The adults must be independent of each other as to whether they are wearing jeans; the probability that each adult is wearing jeans must be the same. (Do not say there must be only two outcomes; this is automatically implied by the question.)

6 More than one relevant outcome on each trial.

7 (a) The boys are not chosen independently of each other.
(b) The probability that a day is warm is not the same for each month.

8 A, B, C and D are not independent events.

Miscellaneous exercise 7 (page 119)

1 0.580

2 (a) $B(20,0.1)$ (b) 0.677

3 (a) $B(8,0.2)$ (b) 0.0563

4 (a) 0.337 (b) 0.135

5 (a) 0.286 (b) 0.754

6 (a) 0.0625, 0.25, 0.375, 0.25, 0.0625
(b) 0.273 (c) 0.313

7 Trials in which the only possible outcomes are 'succeed' and 'fail' are repeated, with the probability of a 'succeed' being the same for all trials, and all trials being independent of one another.
(a) 0.035 (b) 0.138; 83

8 (a) 0.116 (b) 0.386 (c) 0.068

9 For example: probability that each hen lays an egg is the same each day for each hen; or hens lay eggs independently of each other, or independently of whether they laid an egg the previous day.
$\frac{5}{6}$; 1.07, 4.02, 8.04, 6.70

10 (a) 0.9 (b) 0.08 (c) 0.14 (d) 0.3; 0.747

8 Expectation and variance of a random variable

Exercise 8A (page 126)

1 (a) $1\frac{7}{8}$ (b) 0.05

2 4, 3.6

3 $5\frac{1}{9}, 1\frac{35}{81}$

4 (a) $\frac{7}{3}$, 0.745

(b)

y	2	3	4	5	6
$P(Y=y)$	$\frac{1}{36}$	$\frac{1}{9}$	$\frac{5}{18}$	$\frac{1}{3}$	$\frac{1}{4}$

$\frac{14}{3}, \frac{10}{9}$

5 E(A) = \$95 000, E(B) = \$115 000; choose B.

6 (a) 0.3, 0.51 (b) 5.7, 0.51
(c) $E(Y)=6-E(X)$, $Var(Y)=Var(X)$. The distribution of Y (the number of unbroken eggs) is the reflection of the distribution of X in the line $x=3$.

7 $1\frac{25}{36}$, 1.434

8 0.2, 2.8, 1.4

9 $a=b=0.15$, $\sigma=1.7$

10 (a)

x	1	2	3	4
$P(X=x)$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{25}{216}$	$\frac{125}{216}$

(b) 1.172;
0.5177; -20.56

11 (b)

h	0	1	2	3
$P(H=h)$	$\frac{11}{48}$	$\frac{7}{16}$	$\frac{13}{48}$	$\frac{1}{16}$

(c) $E(X)=\frac{7}{3}, E(H)=\frac{7}{6}$ (d) $\frac{13}{18}$

Exercise 8B (page 130)

1 (a) 2.8, 2.408 (b) 0.4550

2 0.84, 0.9073

3 10.5, 6.825

4 0.2581

5 1.6

6 0.1407

Miscellaneous exercise 8 (page 131)

1 (a) 2.56, 1.499 (b) 122.9

2 47.90

3 3, 1.2, 0.3456

4 1.3, 1.01, 3.8, 0.76

(a)

z	3	4	5	6	7	8
$P(Z=z)$	0.15	0.16	0.33	0.19	0.14	0.03

5 Fences at 3.62, 12.38
(a) Yes (b) No (c) No (d) Yes
0.0160, 0.0210

6 (a) 0.8704

(b)

x	1	2	3	4
$P(X=x)$	0.4	0.24	0.144	0.216

(c) 2.176, 1.377

7

w	0	1	2
$P(W=w)$	$\frac{1}{3}$	$\frac{8}{15}$	$\frac{2}{15}$

0.8

8 (a) 0.06 (b) 1.6644

9 (a)

y	0	1	2
$P(Y=y)$	0.006 72	0.069 12	0.248 32

3	4
0.389 12	0.286 72

(b) 2.88, 0.8576

9 The normal distribution

Exercise 9A (page 145)

1 (a) 0.8907 (b) 0.9933 (c) 0.5624 (d) 0.1082
(e) 0.0087 (f) 0.2783 (g) 0.9664 (h) 0.9801
(i) 0.5267 (j) 0.0336 (k) 0.0030 (l) 0.4184
(m) 0.95 (n) 0.05 (o) 0.95 (p) 0.05

2 (a) 0.0366 (b) 0.1202 (c) 0.3417 (d) 0.4394
(e) 0.9555 (f) 0.7804

3 (a) 0.8559 (b) 0.2088 (c) 0.0320 (d) 0.1187
(e) 0.4459 (f) 0.95 (g) 0.98 (h) 0.8064
(i) 0.0164

4 (a) 0.44 (b) 1.165 (c) 2.15 (d) 1.017
(e) 0.24 (f) 1.178 (g) 2.452 (h) 0.758
(i) −2.834 (j) −1.955 (k) −1.035 (l) 0
(m) −2.74 (n) −2.192 (o) −1.677 (p) −0.056
(q) 1.645 (r) 1.282 (s) 2.576 (t) 0.674

Exercise 9B (page 145)

1 (a) 0.9332 (b) 0.0062 (c) 0.7734 (d) 0.0401

2 (a) 0.9522 (b) 0.0098 (c) 0.7475 (d) 0.0038

3 (a) 0.1359 (b) 0.0606 (c) 0.7333 (d) 0.7704
(e) 0.8664

4 0.0668

5 (a) 0.1587 (b) 0.0228

6 (a) 54.35 (b) 40.30 (c) 52.25 (d) 41.85

7 (a) 17.73 (b) 14.62 (c) 17.51 (d) 14.45
(e) 3.29

8 (a) 41.6 (b) 29.9 (c) 37.3 (d) 31.7

9 10.03

10 65

11 42.0, 8.47

12 9.51, 0.303

Exercise 9C (page 148)

1 (a) 0.0370 (b) 0.0062 (c) 0.8209

2 (a) 0.2094 (b) 0.0086 (c) 0.1788 (d) 0.6405
105, 4, 89, 320

3 (a) 13 (b) 84 (c) 121

4 (a) 0.0048 (b) 0.1944 (c) 80, 100

5 1970 hours

6 (a) 0.614 (b) 20.9, 16.3 (c) 17

7 336 ml

8 0.041 kg

9 47 hours^2

10 49.1, 13.4

Exercise 9D (page 157)

1 (a) $np = 10$, $nq = 40$, yes
(b) $np = 6$, $nq = 54$, yes
(c) $np = 0.7$, no
(d) $np = 21$, $nq = 9$, yes
(e) $np = 36$, $nq = 4$, no

2 (a) 0.0288 (b) 0.114 (c) 0.0808 (d) 0.375
$np = 12$, $nq = 28$, valid normal approximation

3 (a) 0.0121 (b) 0.117
$np = 15$, $nq = 285$, valid normal approximation

4 (a) 0.0117 (b) 0.0122, 0.0169
$np = 9$, $nq = 9$, valid normal approximation

5 (a) 0.0083 (b) 0.9110
$np = 6$, $nq = 44$, valid normal approximation

6 0.083
$np = 32$, $np = 8$, valid normal approximation

7 (a) 0.300 (b) 0.301

8 0.0168, 34

9 0.0264, 26

10 (a) 0.189, binomial, $np = 1$
(b) 0.642, binomial, $np = 1$
(c) 0.223, normal, $np = 32.1$, $nq = 17.9$

Miscellaneous exercise 9 (page 158)

1 0.0913

2 0.9213

3 0.960

4 0.933

5 0.957

6 0.525

7 0.367

8 0.936 for Luxibrite, 0.928 for Maxigloss, Luxibrite

9 18.9

10 (a) 0.886 (b) Probably not symmetric

11 (a) 20.2% (b) 81.0 g

12 (a) 70 (b) 45

13 10.3

14 6.82, 0.443

15 (a) 9.68% (b) 28.5% (c) 40

16 (a) 25 (b) 0.6730

17 (b) 0.069

18 0.1977

19 47, 15

20 0.7734

21 (a) 0.488 (b) 281, 5.00

22 $np > 5$ and $n(1-p) > 5$
(a) 0.1853 (b) 0.1838 (c) 0.816%

23 (a) 0.590 (b) 0.081
The model assumes that defective syringes occur independently of each other, but this may not be realistic in a manufacturing process.
It is convenient to use the normal approximation because less calculation is involved; 0.0668.

24 (a) Late trains occur independently of each other. The probability that a train is late is constant.
(b) 0.0184 (c) 0.744
0.803 using the model $\mathrm{N}(22.5, 12.375)$

25 53.18

26 10.2

27 0.397

28 (a) (i) 0.236 (ii) 0.206 (b) 0.672
(c) H, p is closer to $\frac{1}{2}$

29 (a) 46.49% (b) 0.532 m (c) 1.00 m

Revision exercise (page 163)

1 (a) Frequency densities: 1.6, 1.2, 0.833, 0.6
(c) 39.88 mm, 28.2 mm

2 (a) Dice is fair, all six faces are equally likely to turn up.
(b) $2\frac{11}{12}$

3 4.80 minutes
(a) 0.266 (b) 0.149
The distribution is unlikely to be symmetrical because some trains will be very late but no trains will be very early.

4 (a) 0.0148 (c) 10
(d) Slightly surprising; it would happen less than 1 in 20 times.

5 (b) 3.8, 2.6

6 (a) 0.0765 (b) 9.12 (c) 0.000 161

7 (a) $\frac{16}{52}$ or $\frac{4}{13}$
(b) $\mathrm{P}(S=0) = \frac{4}{13}$, $\mathrm{P}(S=2) = \frac{5}{13}$, $\mathrm{P}(S=5) = \frac{4}{13}$
$2\frac{4}{13}$, $3\frac{153}{169}$

8 (a) 0.183 (b) 0.013

9 (a) $\frac{3}{4}$
(b) Assume constant probability of success, independence of outcomes. (i) $\frac{3}{16}$ (ii) $\frac{37}{64}$
(c) The probability of success may vary.

10 The binomial distribution is appropriate when counting the number of successes in a fixed number of trials which are independent of one another, and have a constant probability of success.
$np > 5$, $nq > 5$.
(a) 0.196
(b) 0.933; N(250,125) is the approximation.

11 (b) 16, 7.2
(c) 26 or 27

12 (a) 0.75 (b) 0.28

13 (a) 495 (b) 210 (c) $\frac{14}{33}$

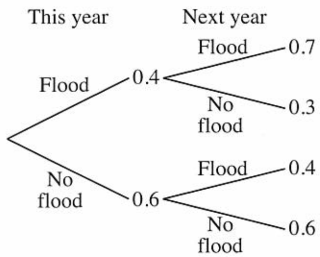
14 104, 38.8

15 0.275; 0.110

Practice examinations

Practice examination 1 (page 167)

1 (i)



(ii) 0.36

2 (ii) $\frac{2555}{1296} \approx 1.97$

3 (i) (a) 29 (b) 42 (c) 21
(ii) Medians are quite similar; men's scores are more spread out.
(iii) The distribution is not symmetrical.

4 (i) 720 (ii) 48 (iii) 144

5 (i) 0.7935 (ii) 89.3

6 (i) (a) 0.258 (b) 0.391 (ii) 0.0135

7 (i) (Frequency densities are 29, 23, 7.5, 4.2, 1.2.)
(ii) $4.985 billion; use of class centres is an approximation.
(iii) Distribution is positively skewed.
(iv) Approximately two-thirds of the companies are worth less than the mean.

Practice examination 2 (page 169)

1 (i) 30 (ii) 13.0

2 (i) 6006 (ii) $\frac{1001}{2584} \approx 0.387$

3 (ii) 0.45, 0.3
(iii) No; e.g. $n = 3$ and $np = 1$ give $p = \frac{1}{3}$, but $\mathrm{P}(X = 3) \neq \frac{1}{27}$.

4 (i) 0.115 (ii) 0.391

5 (i) 0.0705 (ii) 0.815

6 (i) 407 grams, 31 grams
(ii) 408 grams, 20 grams
(iii) With the new variety, weight gains are increased on average, but are more variable.

7 (ii) (a) 33% (b) 8.7 years
(iii) (a) No; the position of the final point doesn't affect the part of the graph where the median is found.
(b) Yes; the centre of the final class interval is affected.

Index

The page numbers given refer to the first mention of each term, or the box if there is one.