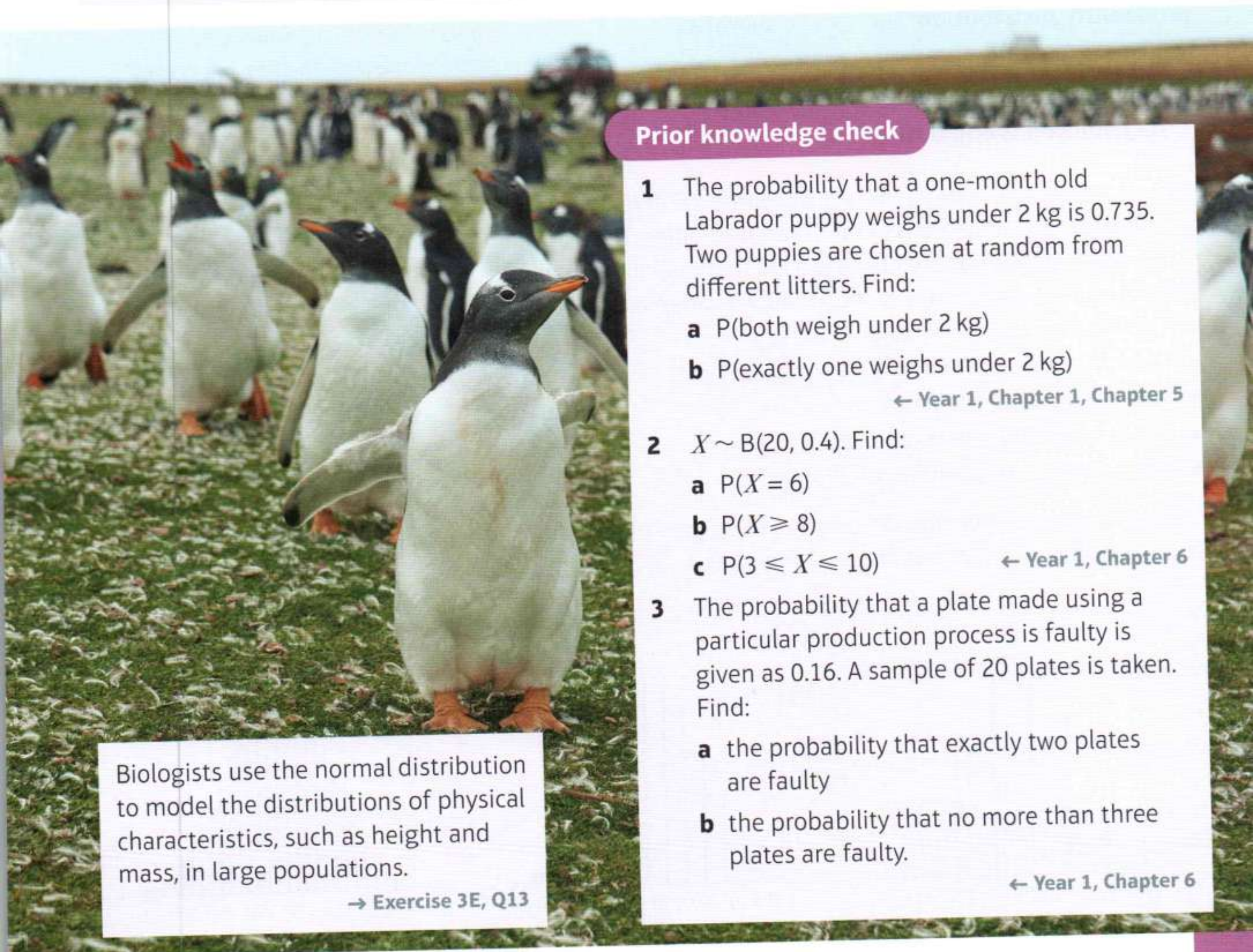
The normal distribution 3

Objectives

After completing this chapter you should be able to:

- Understand the normal distribution and the characteristics of a normal distribution curve → pages 38–41
- Find percentage points on a standard normal curve → pages 41–47
- Calculate values on a standard normal curve → pages 47–49
- Find unknown means and/or standard deviations for a normal distribution → pages 49–53
- Approximate a binomial distribution using a normal distribution → pages 53–55
- Select appropriate distributions and solve real-life problems in context → pages 53–60
- Carry out a hypothesis test for the mean of a normal distribution → pages 53–60



Biologists use the normal distribution to model the distributions of physical characteristics, such as height and mass, in large populations.

→ Exercise 3E, Q13

Prior knowledge check

1 The probability that a one-month old Labrador puppy weighs under 2 kg is 0.735. Two puppies are chosen at random from different litters. Find:

a P(both weigh under 2 kg)

b P(exactly one weighs under 2 kg)

← Year 1, Chapter 1, Chapter 5

2 $X \sim B(20, 0.4)$. Find:

a $P(X = 6)$

b $P(X \geqslant 8)$

c $P(3 \leqslant X \leqslant 10)$

← Year 1, Chapter 6

3 The probability that a plate made using a particular production process is faulty is given as 0.16. A sample of 20 plates is taken. Find:

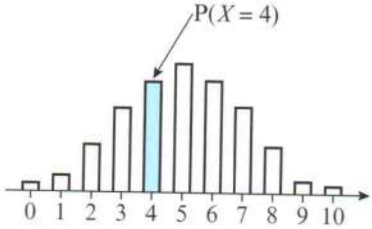
a the probability that exactly two plates are faulty

b the probability that no more than three plates are faulty.

← Year 1, Chapter 6

3.1 The normal distribution

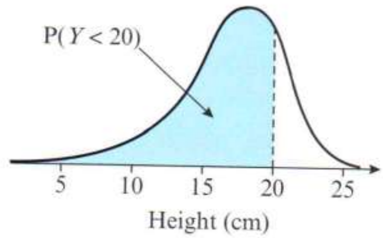
A **continuous random variable** can take any one of infinitely many values. The probability that a continuous random variable takes any one specific value is 0, but you can write the probability that it takes values within a given range. If ten coins are flipped:



X = number of heads

Probability of getting 4 heads is written as $P(X = 4)$

X is a **discrete** random variable



Y = average height of flipped coin

Probability that the average height is less than 20 cm is written as $P(Y < 20)$

Y is a **continuous** random variable

A continuous random variable has a **continuous probability distribution**. This can be shown as a curve on a graph.

- **The area under a continuous probability distribution is equal to 1.**

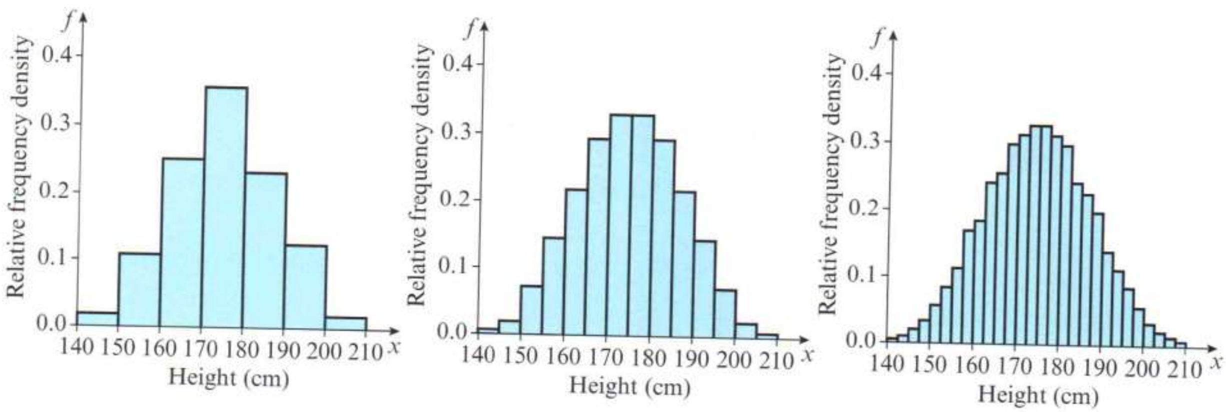
Links A discrete random variable can only take certain distinct values. The sum of the probabilities in a discrete probability distribution is equal to 1.

← Year 1, Chapter 6

The continuous variables generally encountered in real life are more likely to take values grouped around a central value than to take extreme values. The **normal distribution** is a continuous probability distribution that can be used to model many naturally occurring characteristics that behave in this way. Examples of continuous variables that can be modelled using the normal distribution are:

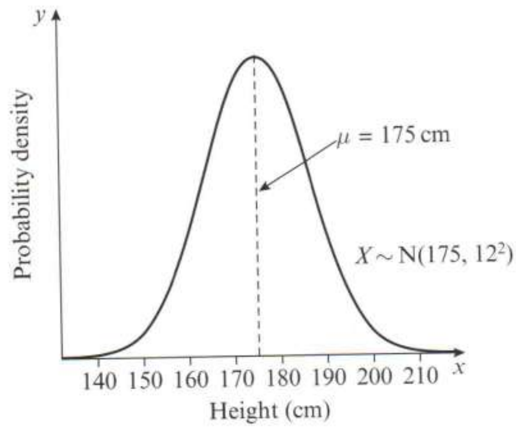
- heights of people within a given population
- weights of tigers in a jungle
- errors in scientific measurements
- size variations in manufactured objects

These histograms show the distribution of heights of adult males in a particular city. As the class width reduces, the distribution gets smoother.



The distribution becomes bell-shaped and is symmetrical about the mean. You can model the heights of adult males in this city using a normal distribution, with mean 175 cm and standard deviation 12 cm.

Notation If X is a normally distributed random variable, you write $X \sim N(\mu, \sigma^2)$ where μ is the population mean and σ^2 is the population variance.



- **The normal distribution**
  - **has parameters μ, the population mean and σ^2, the population variance**
  - **is symmetrical (mean = median = mode)**
  - **has a bell-shaped curve with asymptotes at each end**
  - **has total area under the curve equal to 1**
  - **has points of inflection at $\mu + \sigma$ and $\mu - \sigma$**

For a normally distributed variable:

- approximately 68% of the data lies within one standard deviation of the mean
- 95% of the data lies within two standard deviations of the mean
- nearly all of the data (99.7%) lies within three standard deviations of the mean

Watch out Although a normal random variable could take any value, in practice observations a long way (more than 5 standard deviations) from the mean have probabilities close to 0.

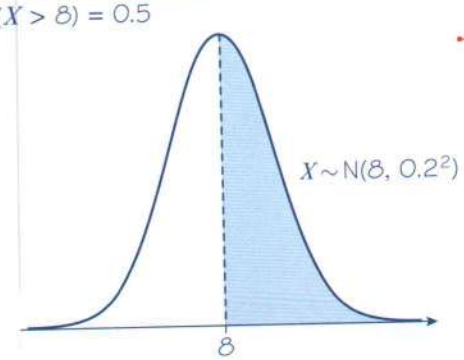
Example 1

The diameters of a rivet produced by a particular machine, X mm, is modelled as $X \sim N(8, 0.2^2)$. Find:

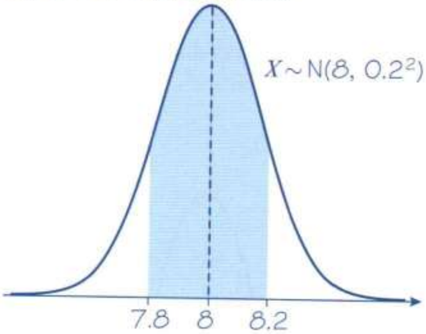
a $P(X > 8)$

b $P(7.8 < X < 8.2)$

a $P(X > 8) = 0.5$



8 is the mean of the distribution. The normal distribution is **symmetrical**, so for any normally distributed random variable $P(X > \mu) = 0.5$.



7.8 and 8.2 are each one standard deviation from the mean. For a normally distributed random variable, 68% of the data lies within one standard deviation of the mean. You can also write $P(\mu - \sigma < X < \mu + \sigma) = 0.68$.

Online Explore the normal distribution curve using technology.

Exercise 3A

1 State, with a reason, whether these random variables are discrete or continuous:

a X, the lengths of a random sample of 100 sidewinder snakes in the Sahara desert

b Y, the scores achieved by 250 students in a university entrance exam

c C, the masses of honey badgers in a random sample of 1000

d Q, the shoe sizes of 200 randomly selected women in a particular town.

2 The lengths, X mm, of a bolt produced by a particular machine are normally distributed with mean 35 mm and standard deviation 0.4 mm. Sketch the distribution of X.

3 The distribution of incomes, in £000s per year, of employees of a bank is shown on the right.
State, with reasons, why the normal distribution is not a suitable model for this data.

4 The armspans of a group of Year 5 pupils, X cm, are modelled as $X \sim N(120, 16)$.

a State the proportion of pupils that have an armspan between 116 cm and 124 cm.

b State the proportion of pupils that have an armspan between 112 cm and 128 cm.

5 The lengths of a colony of adders, Y cm, are modelled as $Y \sim N(100, \sigma^2)$. If 68% of the adders have a length between 93 cm and 107 cm, find σ^2.

(P) **6** The weights of a group of dormice, D grams, are modelled as $D \sim N(\mu, 25)$. If 97.5% of dormice weigh less than 70 grams, find μ.

Problem-solving

Draw a sketch of the distribution. Use the symmetry of the distribution and the fact that 95% of the data lies within 2 standard deviations of the mean.

(P) **7** The masses of the pigs, M kg, on a farm are modelled as $M \sim N(\mu, \sigma^2)$. If 84% of the pigs weigh more than 52 kg and 97.5% of the pigs weigh more than 47.5 kg, find μ and σ^2.

(P) **8** The percentage scores of a group of students in a test, S, are modelled as a normal distribution with mean 45 and standard deviation 15. Find:

a $P(S > 45)$ **b** $P(30 < S < 60)$ **c** $P(15 < S < 75)$

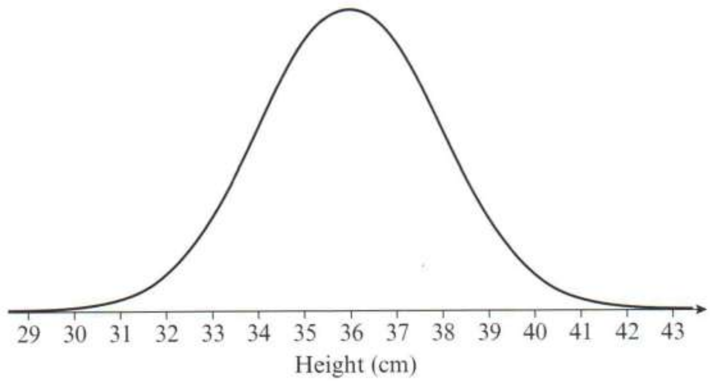
Alexia states that since it is impossible to score above 100%, this is not a suitable model.

d State, with a reason, whether Alexia is correct.

(E) **9** The diagram shows the distribution of heights, in cm, of barn owls in the UK.

An ornithologist notices that the distribution is approximately normal.

Hint The points of inflection on a normal distribution curve occur at $\mu \pm \sigma$.



a State the value of the mean height. **(1 mark)**

b Estimate the standard deviation of the heights. **(2 marks)**

3.2 Finding probabilities for normal distributions

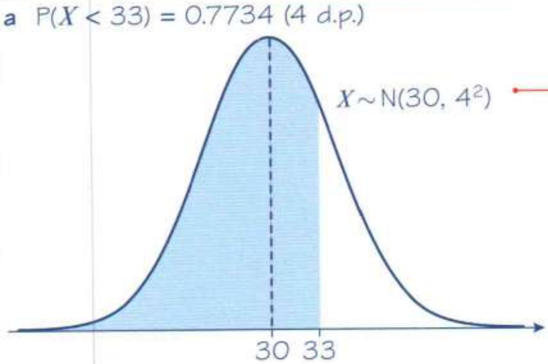
You can find probabilities for a normal distribution using the **normal cumulative distribution** function on your calculator.

Example 2

$X \sim N(30, 4^2)$. Find:

a $P(X < 33)$ **b** $P(X \geqslant 24)$ **c** $P(33.5 < X < 38.2)$ **d** $P(X < 27 \text{ or } X > 32)$

a $P(X < 33) = 0.7734$ (4 d.p.)



You should always draw a sketch to check your answer makes sense. 33 is larger than the mean so the probability should be greater than 0.5.

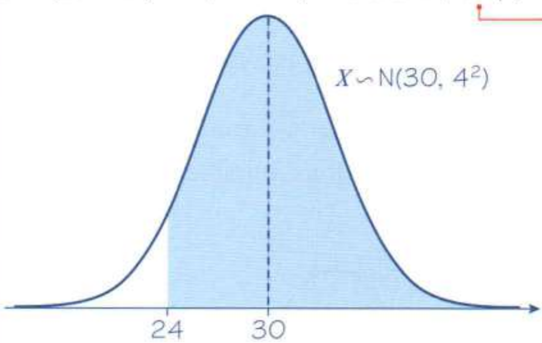
Watch out You need to enter a lower limit into your calculator. Choose a value at least 5 standard deviations away from the mean. For example 10, or −100. Because $P(X < 10)$ is very close to 0, $P(10 < X < 33) \approx P(X < 33)$.

Online Use the Normal CD function on your calculator to find probabilities from a normal distribution.

b $P(X \geqslant 24) = P(X > 24) = 0.9332$ (4 d.p.)

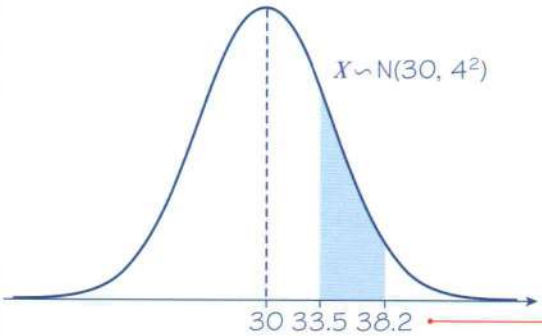
You can use either $>$ or $\geqslant$ interchangeably with a continuous distribution. This is because $P(X = 24) = 0$.

Set the upper limit on your calculator to any large value greater than 5 standard deviations above the mean. You could use 50, 100 or 1000.



c $P(33.5 < X < 38.2) = 0.1706$ (4 d.p.)

Enter both the upper and the lower limits in your calculator.

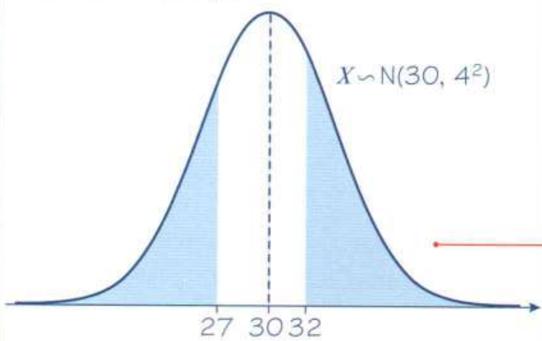


Both 33.5 and 38.2 are above the mean, so the probability should be less than 0.5.

d $P(X < 27 \text{ or } X > 32) = 1 - P(27 < X < 32)$

$= 1 - 0.4648$

$= 0.5352$ (4 d.p.)



Use the fact that the total probability is equal to 1. Sketch the two 'tails' of the required area.

Example 3

An IQ test is applied to a population of adults. The scores, X, on the test are found to be normally distributed with $X \sim N(100, 15^2)$. Adults scoring more than 140 on the test are classified as 'genius'.

a Find the probability that an adult chosen at random achieves a 'genius' classification. Give your answer to three significant figures.

b Twenty adults take the test. Find the probability that two or more are classified as 'genius'.

a $P(X > 140) = 0.00383$ (3 s.f.)

Use your calculator and choose a large upper limit, such as 200.

b Let Y be the number of adults who classify as 'genius'.

$Y \sim B(20, 0.00383)$

This is 20 trials, each with probability of success 0.00383. You can model the number of successful trials as a binomial random variable.

$P(Y \geqslant 2) = 1 - P(Y \leqslant 1)$
$= 1 - 0.9973379...$
$= 0.00266$ (3 s.f.)

Use the binomial cumulative distribution function on your calculator to find $P(Y \leqslant 1)$. You could also find $P(Y = 1) + P(Y = 0)$ which will have the same value. Then subtract this result from 1 to find $P(Y \geqslant 2)$.

← Year 1, Chapter 6

Exercise 3B

1 The random variable $X \sim N(30, 2^2)$.
Find: **a** $P(X < 33)$ **b** $P(X > 26)$ **c** $P(X \geqslant 31.6)$

2 The random variable $X \sim N(40, 9)$.
Find: **a** $P(X > 45)$ **b** $P(X \leqslant 38)$ **c** $P(41 \leqslant X \leqslant 44)$

Watch out In the normal distribution N(40, 9) the second parameter is the **variance**. The standard deviation in this normal distribution is $\sqrt{9} = 3$.

3 The random variable $X \sim N(25, 25)$.
Find: **a** $P(Y < 20)$ **b** $P(18 < Y < 26)$ **c** $P(Y > 23.8)$

4 The random variable $X \sim N(18, 10)$.
Find: **a** $P(X \geqslant 20)$ **b** $P(X < 15)$ **c** $P(18.4 < X < 18.7)$

5 The random variable $M \sim N(15, 1.5^2)$.
a Find: **i** $P(M > 14)$ **ii** $P(M < 14)$
b Calculate the sum of your answers to **a i** and **ii** and comment on your answer.

6 The random variable $T \sim N(4.5, 0.4)$.
a Find $P(T < 4.2)$.
b Without further calculation, write down $P(T > 4.2)$.

(P) **7** The random variable $Y \sim N(45, 2^2)$. Find:
a $P(Y < 41 \text{ or } Y > 47)$ **b** $P(Y < 44 \text{ or } 46.5 < Y < 47.5)$

(E) **8** The volume of soap dispensed by a soap-dispenser on each press, X ml, is modelled as $X \sim N(6, 0.8^2)$.
a Find: **i** $P(X > 7)$ **ii** $P(X < 5)$ **(2 marks)**
The soap dispenser is pressed three times.
b Find the probability that on all three presses, less than 5 ml of soap is dispensed. **(2 marks)**

(E) **9** The amount of mineral water, W ml, in a bottle produced by a certain manufacturer is modelled as $W \sim N(500, 14^2)$.

a Find: **i** $P(W > 505)$ **ii** $P(W < 490)$ **(2 marks)**

A sample of 4 bottles is taken.

b Find the probability that all of the bottles contain more than 490 ml. **(2 marks)**

(P) **10** The heights of a large group of women are normally distributed with a mean of 165 cm and a standard deviation of 3.5 cm. A woman is selected at random from this group.

a Find the probability that she is shorter than 160 cm.

Steven is looking for a woman whose height is between 168 cm and 174 cm for a part in his next film.

b Find the proportion of women from this group who meet Steven's criteria.

A sample of 20 women is taken from the group.

c Find the probability that at least 5 of the women meet Steven's criteria.

Problem-solving

For part **c**, formulate a binomial random variable to represent the number of women in the sample who meet Steven's criteria.

(E/P) **11** The diameters of bolts, D mm, made by a particular machine are modelled as $D \sim N(13, 0.1^2)$.

a Find the probability that a bolt, chosen at random, has a diameter less than 12.8 mm. **(1 mark)**

Bolts are considered to be 'perfect' if the diameter lies between 12.9 mm and 13.1 mm. A random sample of 40 bolts is taken.

b Find the probability that more than 25 of the bolts are 'perfect'. **(4 marks)**

(E/P) **12** The masses, X grams, of a large population of squirrels are modelled as a normal distribution with $X \sim N(480, 40^2)$.

a Find the probability that a squirrel chosen at random has a mass greater than 490 g. **(1 mark)**

A naturalist takes a random sample of 30 squirrels from the population.

b Find the probability that at least 15 of the squirrels have a mass between 470 g and 490 g. **(4 marks)**

3.3 The inverse normal distribution function

For a given probability, p, you can use your calculator to find a value of a such that $P(X < a) = p$. This function is usually called the **inverse normal distribution** function on your calculator.

Example 4

$X \sim N(20, 3^2)$. Find, correct to two decimal places, the values of a such that:

a $P(X < a) = 0.75$

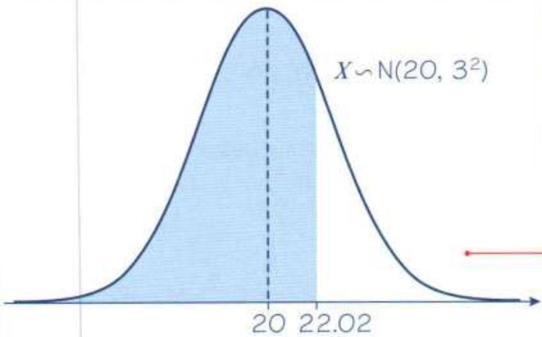
b $P(X > a) = 0.4$

c $P(16 < X < a) = 0.3$

a $a = 22.02$ (2 d.p.)

Enter $\mu = 20$, $\sigma = 3$ and $p = 0.75$ into your calculator. The value for p might be labelled 'Area' on your calculator because it represents the area under the curve to the left of a.

This means that for $X \sim N(20, 3^2)$, $P(X < 22.02) = 0.75$. You can check this result using your calculator.



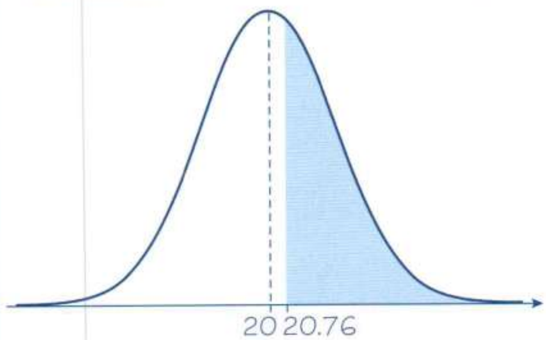
Draw a sketch to check that your answer makes sense. 0.75 is more than 0.5 so the value should be greater than the mean.

b $P(X > a) = 0.4$ so $P(X < a) = 0.6$

$a = 20.76$

Use the fact that $P(X > a) + P(X < a) = 1$ to find the area to the **left** of a before using your calculator.

$P(X < 20.76) = 0.6$, so $P(X > 20.76) = 0.4$ as needed.



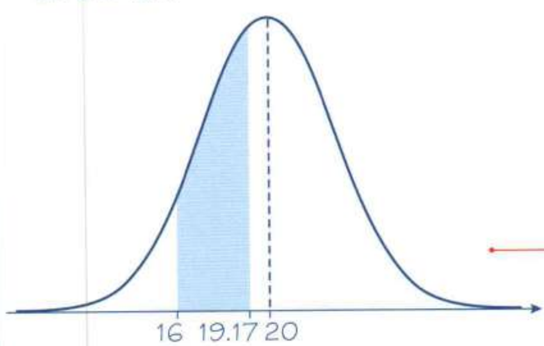
c $P(16 < X < a) = 0.3$

So $P(X < a) = 0.3 + P(X < 16)$

$= 0.3 + 0.09121 = 0.39121$

So $a = 19.17$

Watch out You can't use your calculator to find a directly. Use the fact that $P(16 < X < a) = P(X < a) - P(X < 16)$



You can check your answer using your calculator by working out $P(16 < X < 19.17)$.

Online Use the Inverse Normal function on your calculator to calculate values which satisfy given probability statements for the normal distribution.

Example 5

Plates made using a particular manufacturing process have a diameter, D cm, which can be modelled using a normal distribution, $D \sim N(20, 1.5^2)$.

a Given that 60% of plates are less than x cm, find x.

b Find the interquartile range of the plate diameters.

a $P(D < x) = 0.6$
$\Rightarrow x = 20.38$ cm

b $P(D < Q_1) = 0.25$
$\Rightarrow Q_1 = 18.99$ cm
$P(D < Q_3) = 0.75$
$\Rightarrow Q_3 = 21.01$ cm
The interquartile range is
$21.01 - 18.99 = 2.02$ cm (2 d.p.)

You can use a normal distribution to determine the **proportion** of the data that lie within a certain interval. Use the inverse normal distribution function on your calculator.

25% of the data values lie below the lower quartile, Q_1, and 75% lie below the upper quartile, Q_3. ← Year 1, Chapter 2

The distribution is symmetrical so Q_1 and Q_3 should be the same distance away from the mean.

Exercise 3C

1 The random variable $X \sim N(30, 5^2)$. Find the value of a, to 2 decimal places, such that:
a $P(X < a) = 0.3$ **b** $P(X < a) = 0.75$ **c** $P(X > a) = 0.4$ **d** $P(32 < X < a) = 0.2$

2 The random variable $X \sim N(12, 3^2)$. Find the value of a, to 2 decimal places, such that:
a $P(X < a) = 0.1$ **b** $P(X > a) = 0.65$
c $P(10 \leqslant X \leqslant a) = 0.25$ **d** $P(a < X < 14) = 0.32$

3 The random variable $X \sim N(20, 12)$.
a Find the value of a and the value of b such that:
i $P(X < a) = 0.40$ **ii** $P(X > b) = 0.6915$
b Find $P(b < X < a)$.

4 The random variable $Y \sim N(100, 15^2)$.
a Find the value of a and the value of b such that:
i $P(Y > a) = 0.975$ **ii** $P(Y < b) = 0.10$
b Find $P(a < Y < b)$.

5 The random variable $X \sim N(80, 16)$.
a Find the value of a and the value of b such that:
i $P(X > a) = 0.40$ **ii** $P(X < b) = 0.5636$
b Find $P(b < X < a)$.

(P) **6** The masses, M kg, of a population of badgers are modelled as $M \sim N(4.5, 0.6^2)$.
For this population, find:
a the lower quartile
b the 80th percentile
c Explain without calculation why $Q_2 = 4.5$ kg.

(E) **7** The percentage scores, X, of a group of learner drivers in a theory test is modelled as a normal distribution with $X \sim N(72, 6^2)$.
a Find the value of a such that $P(X < a) = 0.6$. **(1 mark)**
b Find the interquartile range of the scores. **(2 marks)**

E/P **8** The masses, Y grams, of a brand of chocolate bar are modelled as $Y \sim N(60, 2^2)$.

a Find the value of y such that $P(Y > y) = 0.2$. **(1 mark)**

b Find the 10% to 90% interpercentile range of masses. **(2 marks)**

c Tom says that the median is equal to the mean. State, with a reason, whether Tom is correct. **(1 mark)**

E/P **9** The distribution of heights, H cm, of a large group of men is modelled using $H \sim N(170, 10^2)$. A frock coat is a coat that goes from the neck of a person to near the floor. A clothing manufacturer uses the information to make three different lengths of frock coats. The table below shows the proportion of each size they will make.

Short	Regular	Long
30%	50%	20%

a The company wants to advertise a range of heights for which the regular frock coat is suitable. Use the model to suggest suitable heights for the advertisement. **(4 marks)**

b State one assumption you have made in deciding these values. **(1 mark)**

3.4 The standard normal distribution

It is often useful to **standardise** normally distributed random variables. You do this by coding the data so that it can be modelled by the **standard normal distribution**.

- **The standard normal distribution has mean 0 and standard deviation 1.**

Notation The standard normal variable is written as $Z \sim N(0, 1^2)$.

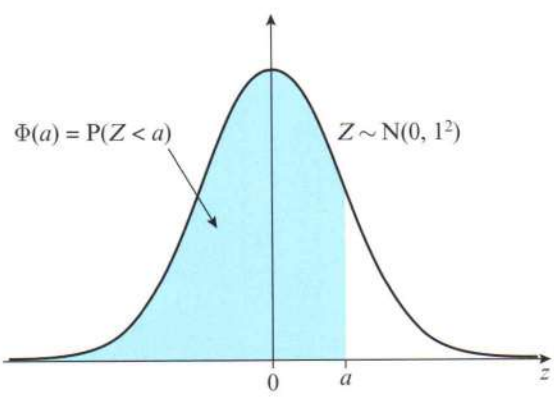
If $X \sim N(\mu, \sigma^2)$ is a normal random variable with mean μ and standard deviation σ, then you can code X using the formula:

$$Z = \frac{X - \mu}{\sigma}$$

Links If $X = x$ then the corresponding value of Z will be $z = \frac{x - \mu}{\sigma}$. The mean of the coded data will be $\frac{\mu - \mu}{\sigma} = 0$ and the standard deviation will be $\frac{\sigma}{\sigma} = 1$. ← Year 1, Section 2.5

The resulting **z-values** will be normally distributed with mean 0 and standard deviation 1.

For the standard normal curve $Z \sim N(0, 1^2)$, the probability $P(Z < a)$ is sometimes written as $\Phi(a)$. You can find it by entering $\mu = 0$ and $\sigma = 1$ into the normal cumulative distribution function on your calculator.



Example 6

The random variable $X \sim N(50, 4^2)$. Write in terms of $\Phi(z)$ for some value z:

a $P(X < 53)$ **b** $P(X \geqslant 55)$

a $z = \frac{53 - 50}{4} = 0.75$

$P(X < 53) = P(Z < 0.75)$
$= \Phi(0.75)$

b $P(X \geqslant 55) = 1 - P(X < 55)$

$z = \frac{55 - 50}{4} = 1.25$

$P(X \geqslant 55) = 1 - P(Z < 1.25)$
$= 1 - \Phi(1.25)$

Code the data so that it is modelled by the standard normal distribution $N(0, 1^2)$.
Use $Z = \frac{X - \mu}{\sigma}$

The distribution is continuous, so you can use $<$ and $\leqslant$ interchangeably.

You sometimes need to find z-values that correspond to given probabilities. You can find these probabilities for some standard values of p by using the percentage points of the normal distribution table on page 191. This table will be given in the *Mathematics Formulae and Statistical Tables* booklet in your exam. It gives values of z and p such that $P(Z > z) = p$.

p	z	p	z
0.5000	0.0000	0.0500	1.6449
0.4000	0.2533	0.0250	1.9600
0.3000	0.5244	0.0100	2.3263
0.2000	0.8416	0.0050	2.5758
0.1500	1.0364	0.0010	3.0902
0.1000	1.2816	0.0005	3.2905

So $P(Z > 1.96) = 0.025$. You can use the symmetry of the distribution to find corresponding negative z-values. $P(Z < -1.96) = 0.025$ so $P(Z > -1.96) = 0.975$.

Example 7

The systolic blood pressure of an adult population, S mmHg, is modelled as a normal distribution with mean 127 and standard deviation 16.

A medical researcher wants to study adults with blood pressures higher than the 95th percentile. Find the minimum blood pressure for an adult included in her study.

$S \sim N(127, 16^2)$

Using the percentage points table:

$P(Z > 1.6449) = 0.05$

$\frac{s - 127}{16} = 1.6449$

$s = 153$ (3 s.f.)

The researcher should include adults with a blood pressure > 153 mmHg

Use the percentage points table with $p = 0.05$.

Convert the value for Z back into a value for S. Remember that the denominator is σ, not σ^2.

You could also find the inverse normal function on your calculator with $\mu = 127$, $\sigma = 16$ and $p = 0.95$.

Exercise 3D

1 For the standard normal distribution $Z \sim N(0, 1^2)$, find:

a $P(Z < 2.12)$ **b** $P(Z < 1.36)$ **c** $P(Z > 0.84)$ **d** $P(Z < -0.38)$
e $P(-2.30 < Z < 0)$ **f** $P(Z < -1.63)$ **g** $P(-2.16 < Z < -0.85)$ **h** $P(-1.57 < Z < 1.57)$

2 For the standard normal distribution $Z \sim N(0, 1^2)$, find values of a such that:

a $P(Z < a) = 0.9082$ **b** $P(Z > a) = 0.0314$
c $P(Z > a) = 0.1500$ **d** $P(Z > a) = 0.9500$
e $P(0 < Z < a) = 0.3554$ **f** $P(0 < Z < a) = 0.4946$
g $P(-a < Z < a) = 0.80$ **h** $P(-a < Z < a) = 0.40$

Hint For parts **g** and **h** you will need to use the symmetry properties of the distribution.

3 The random variable $X \sim N(0.8, 0.05^2)$. For each of the following values of X, write down the corresponding value of the standardised normal distribution, $Z \sim N(0, 1^2)$.

a $x = 0.8$ **b** $x = 0.792$ **c** $x = 0.81$ **d** $x = 0.837$

4 The normal distribution $X \sim N(154, 12^2)$. Write in terms of $\Phi(z)$:

a $P(X < 154)$ **b** $P(X < 160)$
c $P(X > 151)$ **d** $P(140 < X < 155)$

Hint Write your answer to part **d** in the form $\Phi(z_1) - \Phi(z_2)$.

(E) 5 **a** Use the percentage points table to find a value of z such that $P(Z > z) = 0.025$. **(1 mark)**

b A fighter jet training programme takes only the top 2.5% of candidates on a test. Given that the scores can be modelled using a normal distribution with mean 80 and standard deviation 4, use your answer to part **a** to find the score necessary to get on the programme. **(2 marks)**

(E) 6 **a** Use the percentage points table to find a value of z such that $P(Z < z) = 0.15$. **(1 mark)**

b A hat manufacturer makes a special 'petite' hat which should fit 15% of its customers. Given that hat sizes can be modelled using a normal distribution with mean 57 cm and standard deviation 2 cm, use your answer to part **a** to find the size of a 'petite' hat. **(2 marks)**

(E) 7 **a** Use the percentage points table to find the values of z that correspond to the 10% to 90% interpercentile range. **(2 marks)**

A particular brand of light bulb has a life modelled as a normal distribution with mean 1175 hours and standard deviation 56 hours. The bulb life is considered 'standard' if its life falls into the 10% to 90% interpercentile range.

b Use your answer to part **a** to find the range of life to the nearest hour for a 'standard' bulb. **(2 marks)**

3.5 Finding μ and σ

You might need to find an unknown mean or standard deviation for a normally distributed variable.

Example 8

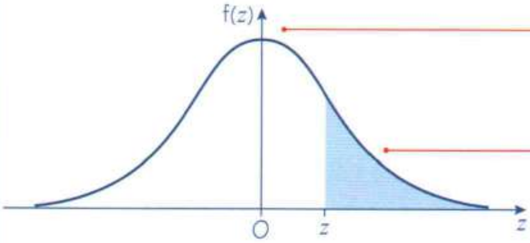
The random variable $X \sim N(\mu, 3^2)$.

Given that $P(X > 20) = 0.20$, find the value of μ.

$P(X > 20) = 0.20$

$P\left(Z > \frac{20 - \mu}{3}\right) = 0.20$

Use $Z = \frac{X - \mu}{\sigma}$



Draw a diagram for Z.

$p = 0.20$

$z = 0.8416$

so $0.8416 = \frac{20 - \mu}{3}$

$\mu = 17.4752 = 17.5$ (3 s.f.)

Problem-solving

You don't know μ, so you need to use the **standard normal distribution**. Use your calculator with $\mu = 0$, $\sigma = 1$ and $p = 0.8$ to find the value of z such that $P(Z > z) = 0.2$. You could also use the percentage points table.

You know one value of X and the corresponding value of Z so use the coding formula to find μ.

Example 9

A machine makes metal sheets with width, X cm, modelled as a normal distribution such that $X \sim N(50, \sigma^2)$.

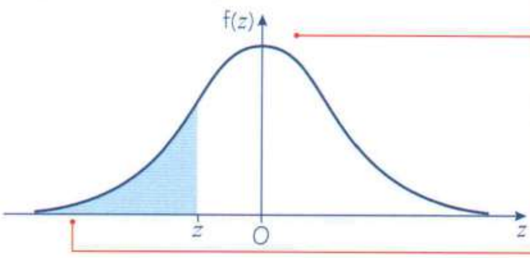
a Given that $P(X < 46) = 0.2119$, find the value of σ.

b Find the 90th percentile of the widths.

a $P(X < 46) = 0.2119$

$P\left(Z < \frac{46 - 50}{\sigma}\right) = 0.2119$

Use $Z = \frac{X - \mu}{\sigma}$



Draw a diagram for Z.

$p = 0.2119$

$P(Z < -0.80) = 0.2119$

So $\frac{46 - 50}{\sigma} = -0.80$

$\frac{-4}{-0.80} = \sigma$

$\sigma = 5$

Use the inverse normal distribution function on your calculator with $\mu = 0$, $\sigma = 1$ and $p = 0.2119$.

Online Use the Inverse Normal function on your calculator with the standard normal distribution.

b $X \sim N(50, 5^2)$.
Let a be the 90th percentile.
$P(X < a) = 0.9$
$a = 56.4$ cm (1 d.p.)

Now that you have calculated σ you can write out the distribution.

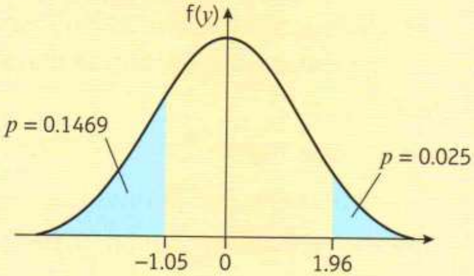
Use the inverse normal distribution function of your calculator with $\mu = 50$, $\sigma = 5$ and $p = 0.9$.

Example 10

The random variable $X \sim N(\mu, \sigma^2)$.
Given that $P(X > 35) = 0.025$ and $P(X < 15) = 0.1469$, find the value of μ and the value of σ.

$P(Z > z_1) = 0.025 \Rightarrow z_1 = 1.96$
$P(Z < z_2) = 0.1469 \Rightarrow z_2 = -1.05$
So $-1.05 = \dfrac{15 - \mu}{\sigma}$
$-1.05\sigma + \mu = 15 \quad (1)$
and $1.96 = \dfrac{35 - \mu}{\sigma}$
$1.96\sigma + \mu = 35 \quad (2)$
(2) − (1): $3.01\sigma = 20$
$\sigma = 6.6445...$
Substituting into (2):
$\mu = 35 - 1.96 \times 6.6445... = 21.976...$
So $\sigma = 6.64$ and $\mu = 22.0$ (3 s.f.)

Find z-values corresponding to a 'right-tail' of 0.025 and a 'left-tail' of 0.1469:



Use $\dfrac{X - \mu}{\sigma}$ to link X and Z values and form two simultaneous equations in μ and σ.

Exercise 3E

1 The random variable $X \sim N(\mu, 5^2)$ and $P(X < 18) = 0.9032$.
Find the value of μ.

2 The random variable $X \sim N(11, \sigma^2)$ and $P(X > 20) = 0.01$.
Find the value of σ.

3 The random variable $Y \sim N(\mu, 40)$ and $P(Y < 25) = 0.15$.
Find the value of μ.

4 The random variable $Y \sim N(50, \sigma^2)$ and $P(Y > 40) = 0.6554$.
Find the value of σ.

(P) 5 The random variable $X \sim N(\mu, \sigma^2)$.
Given that $P(X < 17) = 0.8159$ and $P(X < 25) = 0.9970$, find the value of μ and the value of σ.

(P) 6 The random variable $Y \sim N(\mu, \sigma^2)$.
Given that $P(Y < 25) = 0.10$ and $P(Y > 35) = 0.005$, find the value of μ and the value of σ.

(P) **7** The random variable $X \sim N(\mu, \sigma^2)$.
Given that $P(X > 15) = 0.20$ and $P(X < 9) = 0.20$,
find the value of μ and the value of σ.

Hint Draw a diagram and use symmetry to find μ.

(P) **8** The random variable $X \sim N(\mu, \sigma^2)$.
The lower quartile of X is 25 and the upper quartile of X is 45.
Find the value of μ and the value of σ.

(P) **9** The random variable $X \sim N(0, \sigma^2)$.
Given that $P(-4 < X < 4) = 0.6$, find the value of σ.

(P) **10** The random variable $X \sim N(2.68, \sigma^2)$.
Given that $P(X > 2a) = 0.2$ and $P(X < a) = 0.4$, find the value of σ and the value of a.

(E) **11** An automated pottery wheel is used to make bowls. The diameter of the bowls, D mm, is normally distributed with mean μ and standard deviation 5 mm. Given that 75% of bowls are greater than 200 mm in diameter, find:

a the value of μ **(2 marks)**

b $P(204 < D < 206)$ **(1 mark)**

Three bowls are chosen at random.

c Find the probability that all of the bowls are greater than 205 mm in diameter. **(3 marks)**

(E/P) **12** A loom makes table cloths with an average thickness of 2.5 mm. The thickness, T mm, can be modelled using a normal distribution. Given that 65% of table cloths are less than 2.55 mm thick, find:

a the standard deviation of the thickness **(2 marks)**

b the proportion of table cloths with thickness between 2.4 mm and 2.6 mm. **(1 mark)**

A table cloth can be sold if the thickness is between 2.4 mm and 2.6 mm. A sample of 20 table cloths is taken.

c Find the probability that at least 15 table cloths can be sold. **(3 marks)**

(E/P) **13** The masses of the penguins on an island are found to be normally distributed with mean μ, and standard deviation σ. Given that 10% of the penguins have a mass less than 18 kg and 5% of the penguins have a mass greater than 30 kg,

a sketch a diagram to represent this information **(2 marks)**

b find the value of μ and the value of σ. **(6 marks)**

10 penguins are chosen at random.

c Find the probability that at least 4 of them have a mass greater than 25 kg. **(4 marks)**

(E/P) **14** The length of an adult Dachshund is found to be normally distributed with mean μ and standard deviation σ. Given that 20% of Dachshunds have a length less than 16 inches and 10% have a length greater than 18 inches, find:

a the value of μ and the value of σ **(6 marks)**

b the interquartile range. **(2 marks)**

Challenge

A normally distributed random variable $X \sim N(\mu, \sigma^2)$ has interquartile range q.

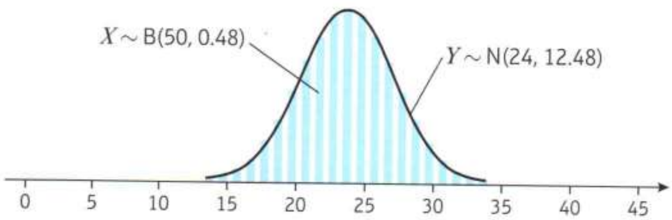
a Show that $\sigma = 0.741q$, where the coefficient of q is correct to 3 s.f.

b Explain why it is not possible to write μ in terms of q only.

3.6 Approximating a binomial distribution

Consider the binomial random variable $X \sim B(n, p)$. It can be difficult to calculate probabilities for X when n is large. In certain circumstances you can use a normal distribution to **approximate** a binomial distribution.

Links The cumulative binomial tables in the formulae booklet only go up to $n = 50$.
← Year 1, Chapter 6



You need to understand the conditions under which this approximation is valid, and learn the relationship between the values of n and p in $B(n, p)$ and the values of μ and σ in the normal approximation $N(\mu, \sigma^2)$.

- **If n is large and p is close to 0.5, then the binomial distribution $X \sim B(n, p)$ can be approximated by the normal distribution $N(\mu, \sigma^2)$ where**
  - $\mu = np$
  - $\sigma = \sqrt{np(1-p)}$

Hint The approximation is only valid when p is close is to 0.5 because the normal distribution is **symmetrical**.

Example 11

A biased coin has P(Head) = 0.53. The coin is tossed 100 times and the number of heads, X, is recorded.

a Write down a binomial model for X.

b Explain why X can be approximated with a normal distribution, $Y \sim N(\mu, \sigma^2)$.

c Find the values of μ and σ in this approximation.

a $X \sim B(100, 0.53)$

b The distribution can be approximated with a normal distribution since n is large and p is close to 0.5.

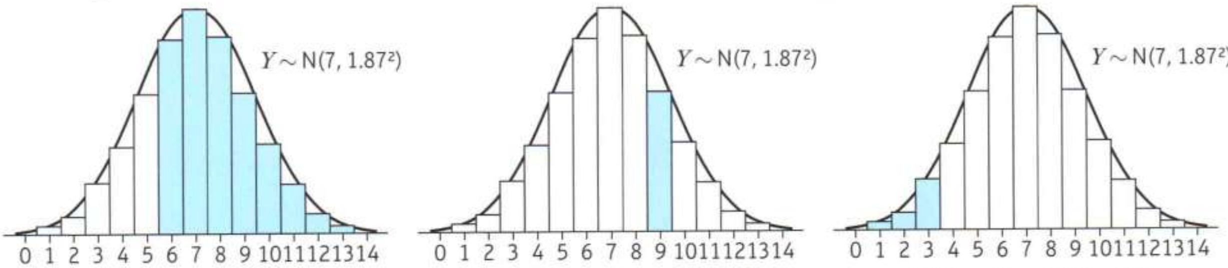
c $\mu = 100 \times 0.53 = 53$ — Use $\mu = np$

$\sigma = \sqrt{100 \times 0.53 \times (1 - 0.53)} = 4.99$ (3 s.f.) — Use $\sigma = \sqrt{np(1-p)}$

The binomial distribution is a discrete distribution but the normal distribution is continuous.

- **If you are using a normal approximation to a binomial distribution, you need to apply a continuity correction when calculating probabilities.**

The diagrams show $X \sim B(14, 0.5)$ being approximated by $Y \sim N(7, 1.87^2)$:



Example 12

The binomial random variable $X \sim B(150, 0.48)$ is approximated by the normal random variable $Y \sim N(72, 6.12^2)$.

Use this approximation to find:

a $P(X \leqslant 70)$ **b** $P(80 \leqslant X < 90)$

Watch out Remember to apply the continuity correction. You are interested in values of the **discrete** random variable X that are less than **or equal to** 70, so you need to consider values less than 70.5 for the continuous random variable Y.

a $P(X \leqslant 70) \approx P(Y < 70.5) = 0.4032$ (4 d.p.)

b $P(80 \leqslant X < 90) \approx P(79.5 < Y < 89.5)$
$= 0.9979 - 0.8898$
$= 0.1081$ (4 d.p.)

For values of X less than 90 consider values of Y less than 89.5.

Example 13

For a particular type of flower bulb, 55% will produce yellow flowers. A random sample of 80 bulbs is planted.

Calculate the percentage error incurred when using a normal approximation to estimate the probability that there are exactly 50 yellow flowers.

Let X = the number of bulbs producing yellow flowers in a sample of 80.

Define a suitable binomial random variable.

Then $X \sim B(80, 0.55)$

$P(X = 50) = \binom{80}{50} 0.55^{50} 0.45^{30} = 0.03651$

Use your calculator to find the exact probability using a binomial distribution. ← Year 1, Chapter 6

X can be approximated by the normal distribution

$Y \sim N(\mu, \sigma^2)$, where $\mu = 80 \times 0.55 = 44$

Use $\mu = np$

$\sigma = \sqrt{80 \times 0.55 \times (1 - 0.55)} = \sqrt{19.8}$

$Y \sim N(44, 19.8)$

Write down the normal approximation.

$P(X = 50) \approx P(49.5 < Y < 50.5)$
$= 0.03618$

To estimate the probability that X takes a single value, apply a continuity correction by considering values half a unit below and half a unit above.

Percentage error $= \dfrac{0.03651 - 0.03618}{0.03651} \times 100$
$= 0.92\%$

Exercise 3F

1 For each of the following binomial random variables, X:

i state, with reasons, whether X can be approximated by a normal distribution.

ii if appropriate, write down the normal approximation to X in the form $N(\mu, \sigma^2)$, giving the values of μ and σ.

a $X \sim B(120, 0.6)$ **b** $X \sim B(6, 0.5)$ **c** $X \sim B(250, 0.52)$

d $X \sim B(100, 0.98)$ **e** $X \sim B(400, 0.48)$ **f** $X \sim B(1000, 0.58)$

2 The random variable $X \sim B(150, 0.45)$. Use a suitable approximation to estimate:

a $P(X \leqslant 60)$ **b** $P(X > 75)$ **c** $P(65 \leqslant X \leqslant 80)$

3 The random variable $X \sim B(200, 0.53)$. Use a suitable approximation to estimate:

a $P(X < 90)$ **b** $P(100 \leqslant X < 110)$ **c** $P(X = 105)$

4 The random variable $X \sim B(100, 0.6)$. Use a suitable approximation to estimate:

a $P(X > 58)$ **b** $P(60 < X \leqslant 72)$ **c** $P(X = 70)$

5 A fair coin is tossed 70 times. Use a suitable approximation to estimate the probability of obtaining more than 45 heads.

6 The probability of a roulette ball landing on red when the wheel is spun is $\frac{50}{101}$.
On one day in a casino, the wheel is spun 1200 times.
Estimate the probability that the ball lands on red in at least half of these spins.

(E) **7** **a** Write down two conditions under which the normal distribution may be used as an approximation to the binomial distribution. **(2 marks)**

A company sells orchids of which 45% produce pink flowers.

A random sample of 20 orchids is taken and X produce pink flowers.

b Find $P(X = 10)$. **(1 mark)**

A second random sample of 240 orchids is taken.

c Using a suitable approximation, find the probability that fewer than 110 orchids produce pink flowers. **(3 marks)**

d The probability that at least q orchids produce pink flowers is 0.2. Find q. **(3 marks)**

(E) **8** A drill bit manufacturer claims that 52% of its bits last longer than 40 hours.
A random sample of 30 bits is taken and X last longer than 40 hours.

a Find $P(X < 17)$. **(1 mark)**

A second random sample of 600 drill bits is taken.

b Using a suitable approximation, find the probability that between 300 and 350 bits (inclusive) last longer than 40 hours. **(3 marks)**

(E/P) **9** A particular breakfast cereal has prizes in 56% of the boxes. A random sample of 100 boxes is taken.

a Find the exact value of the probability that exactly 55 boxes contain a prize. **(1 mark)**

b Find the percentage error when using a normal approximation to calculate the probability that exactly 55 boxes contain prizes. **(4 marks)**

3.7 Hypothesis testing with the normal distribution

You can test hypotheses about the mean of a normally distributed random variable by looking at the mean of a **sample** taken from the whole population.

- **For a random sample of size n taken from a random variable $X \sim N(\mu, \sigma^2)$, the sample mean, $\overline{X}$, is normally distributed with $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.**

Hint If you took lots of different random samples of size n from the population, their means would be normally distributed.

You can use the distribution of the sample mean to determine whether the mean from one particular sample, $\overline{x}$, is statistically significant.

Example 14

A certain company sells fruit juice in cartons. The amount of juice in a carton has a normal distribution with a standard deviation of 3 ml.

The company claims that the mean amount of juice per carton, μ, is 60 ml. A trading inspector has received complaints that the company is overstating the mean amount of juice per carton and he wishes to investigate this complaint. The trading inspector takes a random sample of 16 cartons and finds that the mean amount of juice per carton is 59.1 ml.

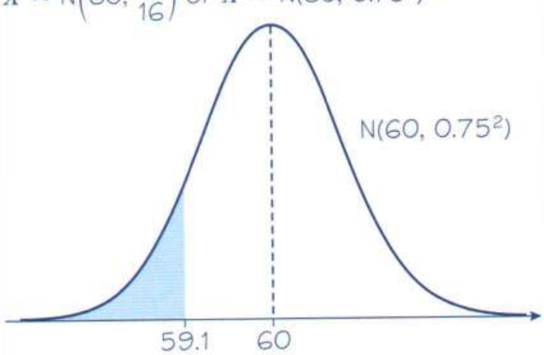
Using a 5% level of significance, and stating your hypotheses clearly, test whether or not there is evidence to uphold this complaint.

$H_0: \mu = 60$

$H_1: \mu < 60$

Let X represent the amount of juice in a carton and assume H_0, so that $X \sim N(60, 3^2)$.

$\overline{X} \sim N\left(60, \frac{3^2}{16}\right)$ or $\overline{X} \sim N(60, 0.75^2)$



$P(\overline{X} < 59.1) = 0.1151$

$0.1151 > 0.05$ so there is insufficient evidence to reject H_0. There is no evidence to uphold the complaint.

The null hypothesis, H_0, is that the population mean is equal to the claimed value.

The inspector is investigating whether the population mean is **less** than 60, so this is a **one-tailed test**.

Write out the population distribution assuming that H_0 is true.

Watch out Your test statistic will be the **sample mean**, $\overline{X}$. This will have the same mean as X, but you need to divide the variance by the sample size. The new variance is $\frac{3^2}{16}$ so the new standard deviation is $\sqrt{\frac{3^2}{16}} = 0.75$.

Use your calculator to find $P(\overline{X} < \overline{x})$.

Compare $P(\overline{X} < 59.1)$ with the significance level of the test. The probability of obtaining this value of $\overline{x}$ is greater than 5%, so you do not reject the null hypothesis. Make sure your conclusion refers to the context given in the problem.

If you need to find a **critical region** or **critical value** for a hypothesis test for the mean of a normal distribution you can standardise your test statistic:

- **For the sample mean of a normally distributed random variable, $\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$, $Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ is a normally distributed random variable with $Z \sim \mathrm{N}(0, 1)$.**

Coding the test statistic in this way allows you to use the percentage points table to determine critical values and critical regions.

You could also use the **inverse normal distribution** function on your calculator to determine critical values and critical regions for $\overline{X}$ directly.

Example 15

A machine produces bolts of diameter D where D has a normal distribution with mean 0.580 cm and standard deviation 0.015 cm.

The machine is serviced and after the service a random sample of 50 bolts from the next production run is taken to see if the mean diameter of the bolts has changed from 0.580 cm. The distribution of the diameters of bolts after the service is still normal with a standard deviation of 0.015 cm.

a Find, at the 1% level, the critical region for this test, stating your hypotheses clearly.

The mean diameter of the sample of 50 bolts is calculated to be 0.587 cm.

b Comment on this observation in light of the critical region.

a $H_0 : \mu = 0.580$, $H_1 : \mu \neq 0.580$

You want to test whether the population mean differs from 0.580 so this is a two-tailed test.

Assume H_0, so that $D \sim \mathrm{N}(0.580, 0.015^2)$

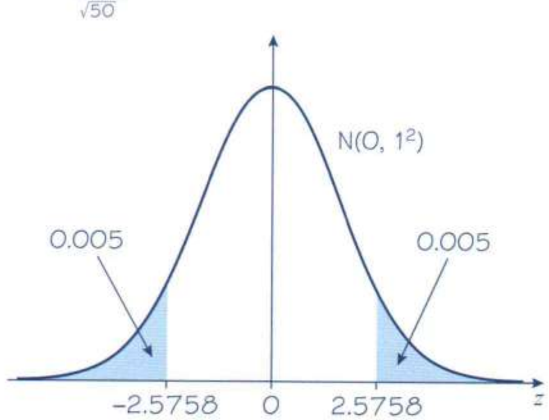
Write out the population distribution assuming that H_0 is true.

$\overline{D} \sim \mathrm{N}\left(0.580, \frac{0.015^2}{50}\right)$

Write out the distribution of the sample means, $\overline{D}$, for a sample of size 50.

Let $Z = \frac{\overline{D} - 0.580}{\frac{0.015}{\sqrt{50}}}$

Standardise $\overline{D}$ using the coding $Z = \frac{\overline{D} - \mu}{\frac{\sigma}{\sqrt{n}}}$. $Z \sim \mathrm{N}(0, 1)$, so you can use the percentage points table for the standard normal distribution.



$\mathrm{P}(Z < z) = 0.005 \Rightarrow z = -2.5758$

$-2.5758 = \frac{\overline{d} - 0.580}{\frac{0.015}{\sqrt{50}}} \Rightarrow \overline{d} = -2.5758 \times \left(\frac{0.015}{\sqrt{50}}\right) + 0.580$

$= 0.5745...$

Use the coding in reverse to find the corresponding value of $\overline{d}$.

$P(Z > z) = 0.005 \Rightarrow z = 2.5758$

$$2.5758 = \frac{\bar{d} - 0.580}{\frac{0.015}{\sqrt{50}}} \Rightarrow \bar{d} = 2.5758 \times \left(\frac{0.015}{\sqrt{50}}\right) + 0.580$$

$$= 0.5854\ldots$$

So the critical region is $\bar{D} \leqslant 0.575$ or $\bar{D} \geqslant 0.585$ (3 s.f.)

b The observed value of $\bar{D}$ (0.587 cm) falls inside the critical region so there is sufficient evidence, at the 1% level, that the mean diameter has changed from 0.580 cm.

Problem-solving

This is a **two-tailed test** with total significance 1%, so you need the probability in each tail to be 0.5% = 0.005. Use the percentage points table to find z-values corresponding to the critical value for each tail.

Remember to refer to the context of the question. Don't just state whether H_0 is rejected or not.

Online Use the inverse normal distribution function on your calculator with $\mu = 0.580$ and $\sigma = \sqrt{\frac{0.015^2}{50}} = 0.002121$ to find the critical region for $\bar{D}$ directly. Set the percentages (or areas) equal to 0.005 and 0.995.

Exercise 3G

1 In each part, a random sample of size n is taken from a population having a normal distribution with mean μ and variance σ^2. Test the hypotheses at the stated levels of significance.

a $H_0: \mu = 21$, $H_1: \mu \neq 21$, $n = 20$, $\bar{x} = 21.2$, $\sigma = 1.5$, at the 5% level

b $H_0: \mu = 100$, $H_1: \mu < 100$, $n = 36$, $\bar{x} = 98.5$, $\sigma = 5.0$, at the 5% level

c $H_0: \mu = 5$, $H_1: \mu \neq 5$, $n = 25$, $\bar{x} = 6.1$, $\sigma = 3.0$, at the 5% level

d $H_0: \mu = 15$, $H_1: \mu > 15$, $n = 40$, $\bar{x} = 16.5$, $\sigma = 3.5$, at the 1% level

e $H_0: \mu = 50$, $H_1: \mu \neq 50$, $n = 60$, $\bar{x} = 48.9$, $\sigma = 4.0$, at the 1% level

2 In each part, a random sample of size n is taken from a population having a $N(\mu, \sigma^2)$ distribution. Find the critical regions for the test statistic $\bar{X}$ in the following tests.

a $H_0: \mu = 120$, $H_1: \mu < 120$, $n = 30$, $\sigma = 2.0$, at the 5% level

b $H_0: \mu = 12.5$, $H_1: \mu > 12.5$, $n = 25$, $\sigma = 1.5$, at the 1% level

c $H_0: \mu = 85$, $H_1: \mu < 85$, $n = 50$, $\sigma = 4.0$, at the 10% level

d $H_0: \mu = 0$, $H_1: \mu \neq 0$, $n = 45$, $\sigma = 3.0$, at the 5% level

e $H_0: \mu = -8$, $H_1: \mu \neq -8$, $n = 20$, $\sigma = 1.2$, at the 1% level

3 The times taken for a capful of stain remover to remove a standard chocolate stain from a baby's bib are normally distributed with a mean of 185 seconds and a standard deviation of 15 seconds. The manufacturers of the stain remover claim to have developed a new formula which will shorten the time taken for a stain to be removed. A random sample of 25 capfuls of the new formula are tested and the mean time for the sample is 179 seconds.

Test, at the 5% level, whether or not there is evidence that the new formula is an improvement.

Hint You are testing for an improvement, so use a **one-tailed** test.

4 The IQ scores of a population are normally distributed with a mean of 100 and a standard deviation of 15. A psychologist wishes to test the theory that eating chocolate before sitting an IQ test improves your score. A random sample of 80 people are selected and they are each given an identical bar of chocolate to eat before taking an IQ test.

a Find, at the 2.5% level, the critical region for this test, stating your hypotheses clearly.

The mean score on the test for the sample of 80 people was 102.5.

b Comment on this observation in light of the critical region.

5 The diameters of circular cardboard drinks mats produced by a certain machine are normally distributed with a mean of 9 cm and a standard deviation of 0.15 cm. After the machine is serviced a random sample of 30 mats is selected and their diameters are measured to see if the mean diameter has altered.

The mean of the sample was 8.95 cm.
Test, at the 5% level, whether there is significant evidence of a change in the mean diameter of mats produced by the machine.

Hint You are testing for an alteration in either direction, so use a **two-tailed** test.

E/P **6** A machine produces metal bolts of diameter D mm, where D is normally distributed with standard deviation 0.1 mm. Bolts with diameter either less than 5.1 mm or greater than 5.6 mm cannot be sold.

Given that 5% of bolts have a diameter in excess of 5.62 mm,

a find the probability that a randomly chosen bolt can be sold. **(5 marks)**

Twelve bolts are chosen.

b Find the probability that fewer than three cannot be sold. **(2 marks)**

A second machine produces bolts of diameter Y mm, where Y is normally distributed with standard deviation 0.08 mm.

A random sample of 20 bolts produced by this machine is taken and the sample mean of the diameters is found to be 5.52 mm.

c Stating your hypotheses clearly, and using a 2.5% level of significance, test whether the mean diameter of all the bolts produced by the machine is less than 5.7 mm. **(4 marks)**

E/P **7** The mass of European water voles, M grams, is normally distributed with standard deviation 12 grams.

Given that 2.5% of water voles have a mass greater than 160 grams,

a find the mean mass of a European water vole. **(3 marks)**

Eight water voles are chosen at random.

b Find the probability that at least 4 have a mass greater than 150 grams. **(3 marks)**

European water rats have mass, N grams, which is normally distributed with standard deviation 85 grams.

A random sample of 15 water rats is taken and the sample mean mass is found to be 875 grams.

c Stating your hypotheses clearly, and using a 10% level of significance, test whether the mean mass of all water rats is different from 860 grams. **(4 marks)**

(E) **8** Daily mean windspeed is modelled as being normally distributed with a standard deviation of 3.1 knots.

A random sample of 25 recorded daily mean windspeeds is taken at Heathrow in 2015.

Given that the mean of the sample is 12.2 knots, test at the 2.5% level of significance whether the mean of the daily mean windspeeds is greater than 9.5 knots.
State your hypotheses clearly. **(4 marks)**

Mixed exercise 3

(E) **1** The heights of a large group of men are normally distributed with a mean of 178 cm and a standard deviation of 4 cm. A man is selected at random from this group.

a Find the probability that he is taller than 185 cm. **(2 marks)**

b Find the probability that three men, selected at random, are all less than 180 cm tall. **(3 marks)**

A manufacturer of door frames wants to ensure that fewer than 1 in 200 men have to stoop to pass through the frame.

c On the basis of this group, find the minimum height of a door frame to the nearest centimetre. **(2 marks)**

(E) **2** The weights of steel sheets produced by a factory are known to be normally distributed with mean 32.5 kg and standard deviation 2.2 kg.

a Find the percentage of sheets that weigh less than 30 kg. **(1 mark)**

Bob requires sheets that weigh between 31.6 kg and 34.8 kg.

b Find the percentage of sheets produced that satisfy Bob's requirements. **(3 marks)**

(E/P) **3** The time a mobile phone battery lasts before needing to be recharged is assumed to be normally distributed with a mean of 48 hours and a standard deviation of 8 hours.

a Find the probability that a battery will last for more than 60 hours. **(2 marks)**

b Find the probability that the battery lasts less than 35 hours. **(1 mark)**

A random sample of 30 phone batteries is taken.

c Find the probability that 3 or fewer last less than 35 hours. **(2 marks)**

(E) **4** The random variable $X \sim N(24, \sigma^2)$. Given that $P(X > 30) = 0.05$, find:

a the value of σ **(2 marks)**

b $P(X < 20)$ **(1 mark)**

c the value of d so that $P(X > d) = 0.01$. **(2 marks)**

(E) **5** A machine dispenses liquid into plastic cups in such a way that the volume of liquid dispensed is normally distributed with a mean of 120 ml. The cups have a capacity of 140 ml and the probability that the machine dispenses too much liquid so that the cup overflows is 0.01.

a Find the standard deviation of the volume of liquid dispensed. **(2 marks)**

b Find the probability that the machine dispenses less than 110 ml. **(1 mark)**

The machine refunds a customer if the liquid in the cup is below a set level.
Given that 10% of customers receive a refund,

c Find the largest volume of liquid that will lead to a refund. **(2 marks)**

E/P **6** The random variable $X \sim N(\mu, \sigma^2)$. The lower quartile of X is 20 and the upper quartile is 40.

a Find μ and σ. **(3 marks)**

b Find the 10% to 90% interpercentile range. **(3 marks)**

E/P **7** The heights of seedlings are normally distributed. Given that 10% of the seedlings are taller than 15 cm and 5% are shorter than 4 cm, find the mean and standard deviation of the heights. **(4 marks)**

E **8** A psychologist gives a student two different tests. The first test has a mean of 80 and a standard deviation of 10 and the student scores 85. Assuming a normal distribution,

a find the probability of scoring 85 or more on the first test. **(2 marks)**

The second test has a mean of 100 and a standard deviation of 15. The student scores 105 on the second test. Assuming a normal distribution,

b find the probability of a score of 105 or more on the second test. **(2 marks)**

c State, giving a reason, which of the student's two test scores was better. **(2 marks)**

E/P **9** Jam is sold in jars and the mean weight of the contents is 108 grams. Only 3% of jars have contents weighing less than 100 grams. Assuming that the weight of jam in a jar is normally distributed, find:

a the standard deviation of the weight of jam in a jar **(2 marks)**

b the proportion of jars where the contents weigh more than 115 grams. **(2 marks)**

A random sample of 25 jars is taken.

c Find the probability that 2 or fewer jars have contents weighing more than 115 grams. **(3 marks)**

E **10** The waiting time at a doctor's surgery is assumed to be normally distributed with standard deviation of 3.8 minutes. Given that the probability of waiting more than 15 minutes is 0.0446, find:

a the mean waiting time **(2 marks)**

b the probability of waiting less than 5 minutes. **(2 marks)**

E/P **11** The thickness of some plastic shelving produced by a factory is normally distributed. As part of the production process the shelving is tested with two gauges. The first gauge is 7 mm thick and 98.61% of the shelving passes through this gauge. The second gauge is 5.2 mm thick and only 1.02% of the shelves pass through this gauge.

Find the mean and standard deviation of the thickness of the shelving. **(4 marks)**

12 A fair coin is spun 60 times. Use a suitable approximation to estimate the probability of obtaining fewer than 25 heads.

E **13** The owner of a local corner shop calculates that the probability of a customer buying a newspaper is 0.40.

A random sample of 100 customers is recorded.

a Give two reasons why a normal approximation may be used in this situation. **(2 marks)**

b Write down the parameters of the normal distribution used. **(2 marks)**

c Use this approximation to estimate the probability that at least half the customers bought a newspaper. **(2 marks)**

(E) **14** The random variable $X \sim B(120, 0.46)$.

a Find $P(X = 65)$. **(1 mark)**

b State why a normal distribution can be used to approximate X, and write down the parameters of such a normal distribution. **(4 marks)**

c Find the percentage error in using the normal approximation to calculate $P(X = 65)$. **(3 marks)**

(E/P) **15** The random variable $Y \sim B(300, 0.6)$.

a Give two reasons why a normal distribution can be used to approximate Y. **(2 marks)**

b Find, using the normal approximation, $P(150 < Y \leqslant 180)$. **(4 marks)**

c Find the largest value of y such that $P(Y < y) < 0.05$. **(3 marks)**

16 Past records from a supermarket show that 40% of people who buy chocolate bars buy the family-size bar. A random sample of 80 people is taken from those who bought chocolate bars. Use a suitable approximation to estimate the probability that more than 30 of these 80 people bought family-size bars.

(E/P) **17** A horticulture company sells apple-tree seedlings. It is claimed that 55% of these seedlings will produce apples within three years.

A random sample of 20 seedlings is taken and X produce apples within three years.

a Find $P(X > 10)$. **(2 marks)**

A second random sample of 200 seedlings is taken. 95 produce apples within three years.

b Assuming the company's claim is correct, use a suitable approximation to find the probability that 95 or fewer seedlings produce apples within three years. **(4 marks)**

c Using your answer to part **b**, comment on the company's claim. **(1 mark)**

(E/P) **18** A herbalist claims that a particular remedy is successful in curing a particular disease in 52% of cases.

A random sample of 25 people who took the remedy is taken.

a Find the probability that more than 12 people in the sample were cured. **(2 marks)**

A second random sample of 300 people was taken and 170 were cured.

b Assuming the herbalist's claim is true, use a suitable approximation to find the probability that at least 170 people were cured. **(4 marks)**

c Using your answer to part **b**, comment on the herbalist's claim. **(1 mark)**

(E) **19** The random variable X has a normal distribution with mean μ and standard deviation 2.

A random sample of 25 observations is taken and the sample mean $\overline{X}$ is calculated in order to test the null hypothesis $\mu = 7$ against the alternative hypothesis $\mu > 7$ using a 5% level of significance.

Find the critical region for $\overline{X}$. **(4 marks)**

(E) **20** A certain brand of mineral water comes in bottles. The amount of water in a bottle, in millilitres, follows a normal distribution of mean μ and standard deviation 2. The manufacturer claims that μ is 125. In order to maintain standards the manufacturer takes a sample of 15 bottles and calculates the mean amount of water per bottle to be 124.2 millilitres.

Test, at the 5% level, whether or not there is evidence that the value of μ is lower than the manufacturer's claim. State your hypotheses clearly. **(4 marks)**

(E) **21** Climbing rope produced by a manufacturer is known to be such that one-metre lengths have breaking strengths that are normally distributed with mean 170.2 kg and standard deviation 10.5 kg. Find, to 3 decimal places, the probability that:

a a one-metre length of rope chosen at random from those produced by the manufacturer will have a breaking strength of 175 kg to the nearest kg **(2 marks)**

b a random sample of 50 one-metre lengths will have a mean breaking strength of more than 172.4 kg. **(3 marks)**

A new component material is added to the ropes being produced. The manufacturer believes that this will increase the mean breaking strength without changing the standard deviation. A random sample of 50 one-metre lengths of the new rope is found to have a mean breaking strength of 172.4 kg.

c Perform a significance test at the 5% level to decide whether this result provides sufficient evidence to confirm the manufacturer's belief that the mean breaking strength is increased. State clearly the null and alternative hypotheses that you are using. **(3 marks)**

/P **22** A machine fills 1 kg packets of sugar. The actual weight of sugar delivered to each packet can be assumed to be normally distributed. The manufacturer requires that,

i the mean weight of the contents of a packet is 1010 g, and

ii 95% of all packets filled by the machine contain between 1000 g and 1020 g of sugar.

a Show that this is equivalent to demanding that the variance of the sampling distribution, to 2 decimal places, is equal to 26.03 g^2. **(3 marks)**

A sample of 8 packets was selected at random from those filled by the machine. The weights, in grams, of the contents of these packets were

1012.6 1017.7 1015.2 1015.7 1020.9 1005.7 1009.9 1011.4

Assuming that the variance of the actual weights is 26.03 g^2,

b test at the 2% significance level (stating clearly the null and alternative hypotheses that you are using) to decide whether this sample provides sufficient evidence to conclude that the machine is not fulfilling condition **i**. **(4 marks)**

/P **23** The diameters of eggs of the little-gull are approximately normally distributed with mean 4.11 cm and standard deviation 0.19 cm.

a Calculate the probability that an egg chosen at random has a diameter between 3.9 cm and 4.5 cm. **(3 marks)**

A sample of 8 little-gull eggs was collected from a particular island and their diameters, in cm, were

4.4, 4.5, 4.1, 3.9, 4.4, 4.6, 4.5, 4.1

b Assuming that the standard deviation of the diameters of eggs from the island is also 0.19 cm, test, at the 1% level, whether the results indicate that the mean diameter of little-gull eggs on this island is different from elsewhere. **(4 marks)**

P **24** The random variable X is normally distributed with mean μ and variance σ^2.

a Write down the distribution of the sample mean $\overline{X}$ of a random sample of size n. **(1 mark)**

A construction company wishes to determine the mean time taken to drill a fixed number of holes in a metal sheet.

b Determine how large a random sample is needed so that the expert can be 95% certain that the sample mean time will differ from the true mean time by less than 15 seconds. Assume that it is known from previous studies that $\sigma = 40$ seconds. **(4 marks)**

Challenge

A football manager claims to have the support of 48% of all the club's fans.

A random sample of 15 fans is taken.

a Find the probability that more than 8 of these fans supported the manager.

A second random sample of 250 fans was taken, and is used to test the football manager's claim at the 5% significance level.

b Use a suitable approximation to find the critical regions for this test.

It was found that 102 fans said they supported the manager.

c Using your answer to part **b**, comment on the manager's claim.

Summary of key points

1 The area under a continuous probability distribution is equal to 1.

2 If X is a normally distributed random variable, you write $X \sim \text{N}(\mu, \sigma^2)$ where μ is the population mean and σ^2 is the population variance.

3 The normal distribution

- has parameters μ, the population mean, and σ^2, the population variance
- is symmetrical (mean = median = mode)
- has a bell-shaped curve with asymptotes at each end
- has total area under the curve equal to 1
- has points of inflection at $\mu + \sigma$ and $\mu - \sigma$

4 The standard normal distribution has mean 0 and standard deviation 1.
The standard normal variable is written as $Z \sim \text{N}(0, 1^2)$.

5 If n is large and p is close to 0.5, then the binomial distribution $X \sim \text{B}(n, p)$ can be approximated by the normal distribution $\text{N}(\mu, \sigma^2)$ where

- $\mu = np$
- $\sigma = \sqrt{np(1-p)}$

6 If you are using a normal approximation to a binomial distribution, you need to apply a **continuity correction** when calculating probabilities.

7 For a random sample of size n taken from a random variable $X \sim \text{N}(\mu, \sigma^2)$, the sample mean, $\overline{X}$, is normally distributed with $\overline{X} \sim \text{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

8 For the sample mean of a normally distributed random variable, $\overline{X} \sim \text{N}\left(\mu, \frac{\sigma^2}{n}\right)$,
$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ is a normally distributed random variable with $Z \sim \text{N}(0, 1)$.