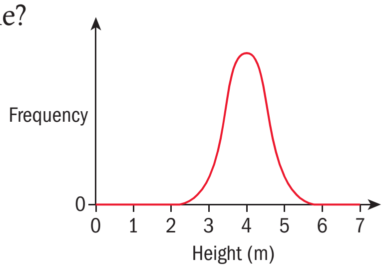
5.1 The normal distribution

For his Mathematical Studies Project, Pedro measures the heights of all the apple trees in his father's orchard. There are 150 trees.

If Pedro drew a diagram to represent the frequency of the heights of all 150 trees, what do you think it would look like?

Pedro then measures the heights of the apple trees in his uncle's orchard. If he drew a diagram of the frequencies of these heights, do you think that this diagram would look different to the previous one?

In both orchards there would probably be a few very small trees and a few very large trees – but those would be the exception. Most of the trees would fall within a certain range of heights. They would roughly fit a bell-shaped curve that is symmetrical about the mean. We call this a **normal distribution**.



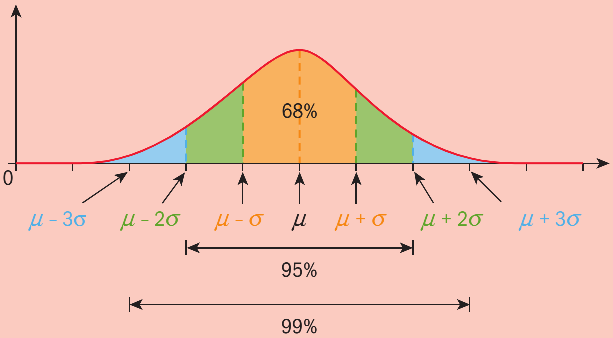
▲ Normal distribution diagram for the tree heights measured by Pedro

Many events fit this type of distribution: for example, the heights of 21-year-old males, the results of a national mathematics examination, the weights of newborn babies, etc.

The properties of a normal distribution

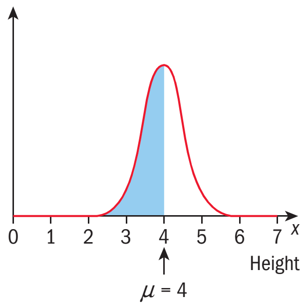
→ The **normal distribution** is the most important continuous distribution in statistics. It has these properties:

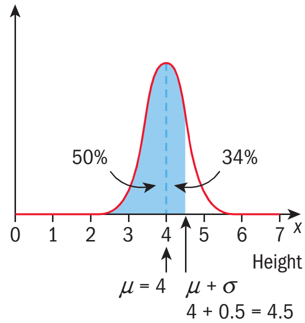
- It is a bell-shaped curve.
- It is symmetrical about the mean, μ. (The mean, the mode and the median all have the same value.)
- The x-axis is an asymptote to the curve.
- The total area under the curve is 1 (or 100%).
- 50% of the area is to the left of the mean and 50% to the right.
- Approximately 68% of the area is within 1 standard deviation, σ, of the mean.
- Approximately 95% of the area is within 2 standard deviations of the mean.
- Approximately 99% of the area is within 3 standard deviations of the mean.



The normal curve is sometimes called the 'Gaussian curve' after the German mathematician Carl – Friedrich Gauss (1777–1855). Gauss used the normal curve to analyze astronomical data in 1809. A portrait of Gauss and the normal curve appear on the old German 10 Deutschmark note.

You can calculate the probabilities of events that follow a normal distribution.
Returning to Pedro and the apple trees, imagine that the mean height of the trees is 4 m and the standard deviation is 0.5 m.
Let the height of an apple tree be x.





From the properties of the normal distribution: Area to left of μ = 50%. Area between μ and $\mu + \sigma$ = 34% (68% ÷ 2).

The probability that an apple tree is less than 4 m is $P(x < 4) = 50\%$ or 0.5. And $P(x < 4.5) = 50\% + 34\% = 84\%$ or 0.84.

→ The **expected value** is found by multiplying the number in the sample by the probability.

For example, if we chose 100 apple trees at random, the expected number of trees that would be less than 4 m = 100 × 0.5 = 50.

Example 1

The waiting times for an elevator are normally distributed with a mean of 1.5 minutes and a standard deviation of 20 seconds.

a Sketch a normal distribution diagram to illustrate this information, indicating clearly the mean and the times within one, two and three standard deviations of the mean.

b Find the probability that a person waits longer than 2 minutes 10 seconds for the elevator.

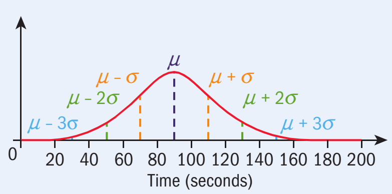
c Find the probability that a person waits less than 1 minute 10 seconds for the elevator.

200 people are observed and the length of time they wait for an elevator is noted.

d Calculate the number of people expected to wait less than 50 seconds for the elevator.

Answers

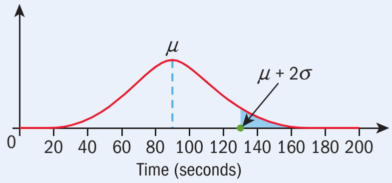
a



1.5 minutes = 90 seconds
μ = mean = 90 seconds
σ = standard deviation = 20 seconds

▶ Continued on next page

b



P(waiting longer than 2 minutes 10 seconds) = 2.5%, or 0.025.

2 minutes 10 seconds = 130 seconds

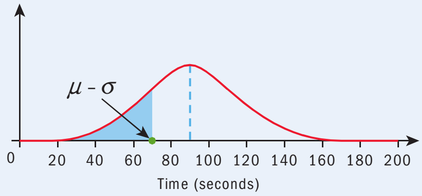
Using symmetry about μ:

Area to right of μ = 50%

Area between μ and $\mu + 2\sigma$ = 47.5% (95% ÷ 2)

Area to right of $\mu + 2\sigma$ = 50% − 47.5% = 2.5%

c



P(waiting less than 1 minute 10 seconds) = 16%, or 0.16.

1 minute 10 seconds = 70 seconds

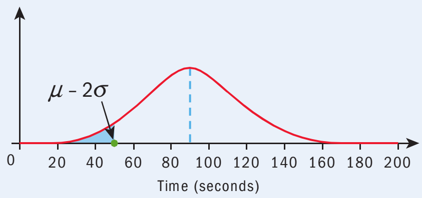
Using symmetry about μ:

Area to left of μ = 50%

Area between μ and $\mu - \sigma$ = 34% (68% ÷ 2)

Area to left of $\mu - \sigma$ = 50% − 34% = 16%

d



P(waiting less than 50 seconds) = 2.5%, or 0.025

So, the expected number of people = 200 × 0.025 = 5.

First find the probability of waiting less than 50 seconds.

Using symmetry about μ:

Area to left of μ = 50%

Area between μ and $\mu - 2\sigma$ = 47.5% (95% ÷ 2)

Area to left of $\mu - 2\sigma$ = 50% − 47.5% = 2.5%

There are 200 people in the sample.

Example 2

The heights of 250 twenty-year-old women are normally distributed with a mean of 1.68 m and standard deviation of 0.06 m.

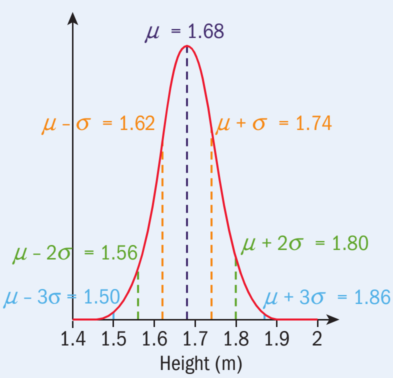
a Sketch a normal distribution diagram to illustrate this information, indicating clearly the mean and the heights within one, two and three standard deviations of the mean.

b Find the probability that a woman has a height between 1.56 m and 1.74 m.

c Find the expected number of women with a height greater than 1.8 m.

Answers

a



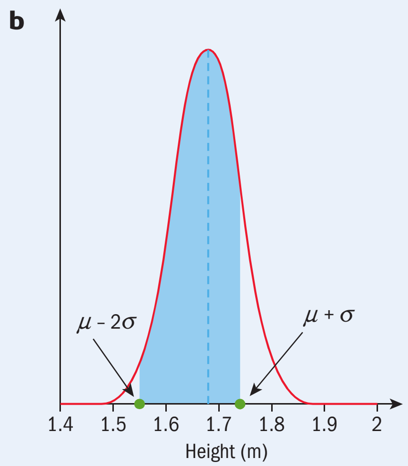
Let

μ = mean = 1.68 m

σ = standard deviation = 0.06 m

Continued on next page

b



P(height between 1.56 m and 1.74 m)
= 81.5%, or 0.815.

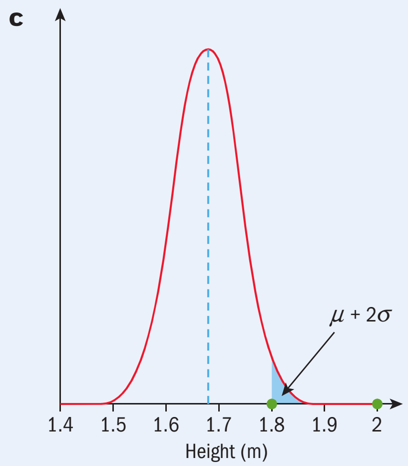
Using symmetry about μ:

Area between μ and μ + σ = 34% (68% ÷ 2)

Area between μ and μ − 2σ = 47.5% (95% ÷ 2)

Area between 1.56 m and 1.74 m = 34% + 47.5%
= 81.5%

c



P(height greater than 1.8 m) = 2.5%, or 0.025.
So, the expected number of women
= 250 × 0.025 = 6.25, or 6 women.

First find the probability of a woman being taller than 1.8 m.

Using symmetry about μ:

Area to right of μ = 50%

Area between μ and μ + 2σ = 47.5% (95% ÷ 2)

Area to right of μ + 2σ = 50% − 47.5% = 2.5%

There are 250 women in the sample.

Exercise 5A

EXAM-STYLE QUESTION

1 The heights of 200 lilies are normally distributed with a mean of 40 cm and a standard deviation of 3 cm.

a Sketch a normal distribution diagram to illustrate this information. Indicate clearly the mean and the heights within one, two and three standard deviations of the mean.

b Find the probability that a lily has a height less than 37 cm.

c Find the probability that a lily has a height between 37 cm and 46 cm.

d Find the expected number of lilies with a height greater than 43 cm.

EXAM-STYLE QUESTIONS

2 100 people were asked to estimate the length of one minute. Their estimates were normally distributed with a mean time of 60 seconds and a standard deviation of 4 seconds.

a Sketch a normal distribution diagram to illustrate this information, indicating clearly the mean and the times within one, two and three standard deviations of the mean.

b Find the percentage of people who estimated between 52 and 64 seconds.

c Find the expected number of people estimating less than 60 seconds.

3 60 students were asked how long it took them to travel to school. Their travel times are normally distributed with a mean of 20 minutes and a standard deviation of 5 minutes.

a Sketch a normal distribution diagram to illustrate this information, indicating clearly the mean and the times within one, two and three standard deviations of the mean.

b Find the percentage of students who took longer than 25 minutes to travel to school.

c Find the expected number of students who took between 15 and 25 minutes to travel to school.

4 Packets of coconut milk are advertised to contain 250 ml. Akshat tests 75 packets. He finds that the contents are normally distributed with a mean volume of 255 ml and a standard deviation of 8 ml.

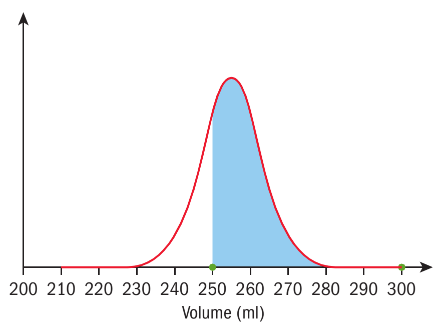
a Sketch a normal distribution diagram to illustrate this information, indicating clearly the mean and the volumes within one, two and three standard deviations of the mean.

b Find the probability that a packet contains less than 239 ml.

c Find the expected number of packets that contain more than 247 ml.

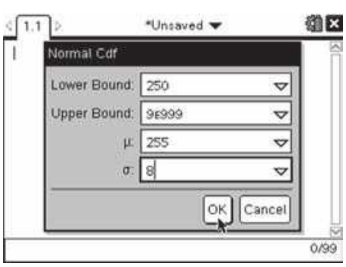
You can use your GDC to calculate values that are not whole multiples of the standard deviation.

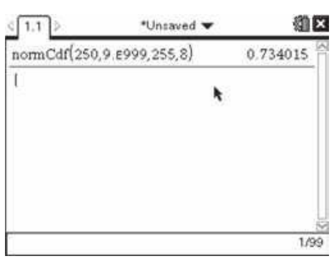
For example, in question 4 of Exercise 5A, suppose we wanted to find the probability that a packet contains more than 250 ml.
First sketch a normal distribution diagram.



In a Calculator page press MENU 5:Probability | 5:Distributions | 2:Normal Cdf and enter the lower bound (250), the upper bound (9×10^{999} – a *very* large number), the mean (255) and the standard deviation (8) in the wizard.

To enter 9×10^{999} you need to type 9E999, but you cannot use the E key. Instead, you must use the EE key.





GDC help on CD: *Alternative demonstrations for the TI-84 Plus and Casio FX-9860GII GDCs are on the CD.*

So, 73.4% of the packets contain more than 250 ml of coconut milk. Alternatively, enter normCdf, the lowest value, the highest value, the mean and the standard deviation directly into the calculator screen.

For a very small number enter -9×10^{999}

Example 3

The lifetime of a light bulb is normally distributed with a mean of 2800 hours and a standard deviation of 450 hours.

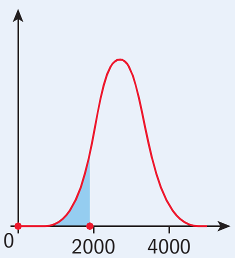
a Find the percentage of light bulbs that have a lifetime of less than 1950 hours.
b Find the percentage of light bulbs that have a lifetime between 2300 and 3500 hours.
c Find the probability that a light bulb has a lifetime of more than 3800 hours.

120 light bulbs are tested.

d Find the expected number of light bulbs with a lifetime of less than 2000 hours.

Answers

a



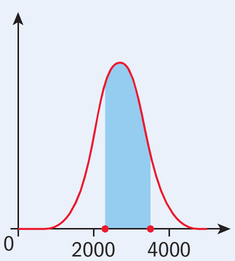
2.95% of the light bulbs have a lifetime of less than 1950 hours.

μ = mean = 2800 hours
σ = standard deviation = 450 hours

Lifetime less than 1950 hours:
lower bound = -9×10^{999}
upper bound = 1950

From GDC:
normCdf(−9E999, 1950, 2800, 450) = 0.029 45
= 2.95%

b



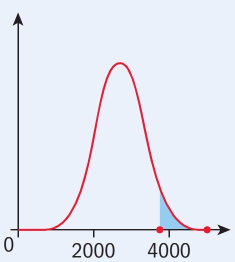
80.7% of the light bulbs have a lifetime between 2300 and 3500 hours.

Lifetime between 2300 and 3500 hours:
lower bound = 2300
upper bound = 3500

Remember not to use −9E999 notations in an exam.

From GDC:
normCdf(2300, 3500, 2800, 450) = 0.8068 = 80.7%

c



Only 1.31% of the light bulbs have a lifetime of more than 3800 hours.

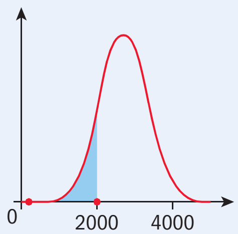
Lifetime more than 3800 hours:
lower bound = 3800
upper bound = 9×10^{999}

GDC help on CD: *Alternative demonstrations for the TI-84 Plus and Casio FX-9860GII GDCs are on the CD.*

From GDC:
normCdf(3800, 9E999, 2800, 450) = 0.0131 = 1.31%

Continued on next page

d



P(lifetime less than 2000 hours) = 3.77%
Expected number = 120 × 0.0377
= 4.524
So, you would expect 4 or 5 light bulbs to have a lifetime of less than 2000 hours.

First find P(lifetime less than 2000 hours):
lower bound = -9×10^{999}
upper bound = 2000

From GDC:
normCdf(−9E999, 2000, 2800, 450) = 0.0377 = 3.77%
120 light bulbs are tested.

Exercise 5B

EXAM-STYLE QUESTION

1 Jordi delivers daily papers to a number of homes in a village. The time taken to deliver the papers follows a normal distribution with mean 80 minutes and standard deviation 7 minutes.

a Sketch a normal distribution diagram to illustrate this information.

b Find the probability that Jordi takes longer than 90 minutes to deliver the papers.

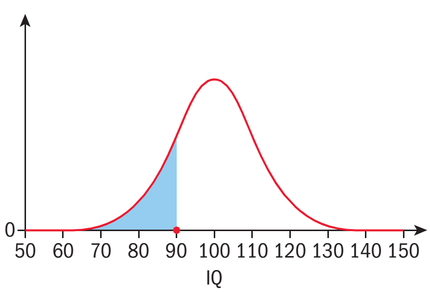
Jordi delivers papers every day of the year (365 days).

c Calculate the expected number of days on which it would take Jordi longer than 90 minutes to deliver the papers.

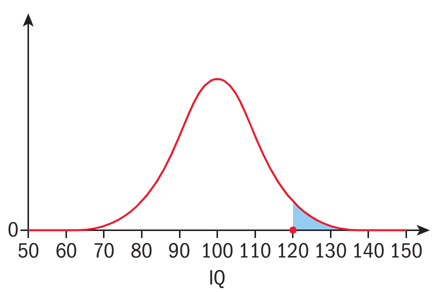
2 A set of 2000 IQ scores is normally distributed with a mean of 100 and a standard deviation of 10.

a Calculate the probability that is represented by each of the following diagrams.

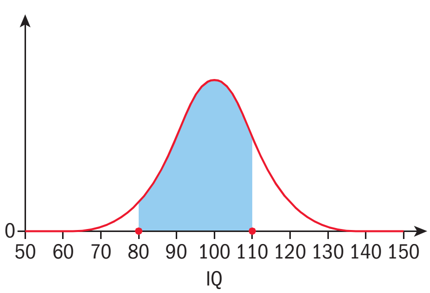
i



ii



iii



Lambert Quételet (1796–1874), a Flemish scientist, was the first to apply the normal distribution to human characteristics. He noticed that measures such as height, weight and IQ were normally distributed.

b Find the expected number of people with an IQ of more than 115.

3 A machine produces washers whose diameters are normally distributed with a mean of 40 mm and a standard deviation of 2 mm.

a Find the probability that a washer has a diameter less than 37 mm.

b Find the probability that a washer has a diameter greater than 45 mm.

Every week 300 washers are tested.

c Calculate the expected number of washers that have a diameter between 35 mm and 43 mm.

EXAM-STYLE QUESTIONS

4 In a certain school, the monthly incomes of members of staff are normally distributed with a mean of 2500 euros and a standard deviation of 400 euros.

a Sketch a normal distribution diagram to illustrate this information.

b Find the probability that a member of staff earns less than 1800 euros per month.

The school has 80 members of staff.

c Calculate the expected number of staff who earn more than 3400 euros.

5 The lengths of courgettes are normally distributed with a mean of 16 cm and a standard deviation of 0.8 cm.

a Find the percentage of courgettes that have a length between 15 cm and 17 cm.

b Find the probability that a courgette is longer than 18 cm.

The lengths of 100 courgettes are measured.

c Calculate the expected number of courgettes that have a length less than 14.5 cm.

6 At a market, the weights of bags of kiwi fruit are normally distributed with a mean of 500 g and a standard deviation of 8 g.
A man picks up a bag of kiwi fruit at random.
Find the probability that the bag weighs more than 510 g.

EXAM-STYLE QUESTIONS

7 The scores in a Physics test follow a normal distribution with mean 70% and standard deviation 8%.

a Find the percentage of students who scored between 55% and 80%.

30 students took the physics test.

b Calculate the expected number of students who scored more than 85%.

8 A machine produces pipes such that the length of each pipe is normally distributed with a mean of 1.78 m and a standard deviation of 2 cm.
Any pipe whose length is greater than 1.83 m is rejected.

a Find the probability that a pipe will be rejected.

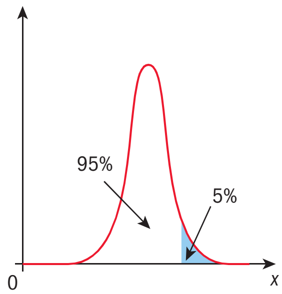
500 pipes are tested.

b Calculate the expected number of pipes that will be rejected.

Inverse normal calculations

Sometimes you are given the percentage area under the curve, i.e. the probability or proportion, and are asked to find the value corresponding to it. This is called an inverse normal calculation.

Always make a sketch to illustrate the information given.
You must always remember to use the area to the **left** when using your GDC. If you are given the area to the **right** of the value, you must subtract this from 1 (or 100%) before using your GDC.
For example, an area of 5% above a certain value means there is an area of 95% below it.

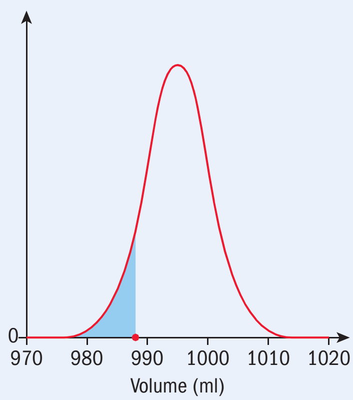


In examinations, inverse normal questions will not involve finding the mean or standard deviation.

Example 4

The volume of cartons of milk is normally distributed with a mean of 995 ml and a standard deviation of 5 ml.
It is known that 10% of the cartons have a volume less than x ml.
Find the value of x.

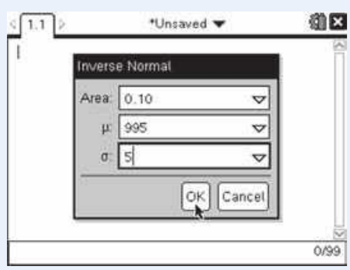
Answer

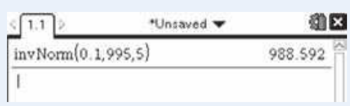


$x = 989$ ml (to 3 sf)

First sketch a diagram. The shaded area represents 10% of the cartons.
Using the GDC:

In a Calculator page press MENU 5:Probability | 5:Distributions | 3:Inverse Normal...
Enter the percentage given (as a decimal, 0.1), the mean (995) and the standard deviation (5).





$x = 989$ (3 sf)
$x = 989$ ml means that 10% of the cartons have a volume less than 989 ml.

GDC help on CD: *Alternative demonstrations for the TI-84 Plus and Casio FX-9860GII GDCs are on the CD.*

Example 5

The weights of pears are normally distributed with a mean of 110 g and a standard deviation of 8 g.

a Find the percentage of pears that weigh between 100 g and 130 g.

It is known that 8% of the pears weigh more than *m* g.

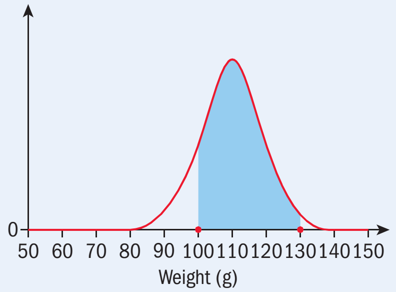
b Find the value of *m*.

250 pears are weighed.

c Calculate the expected number of pears that weigh less than 105 g.

Answers

a



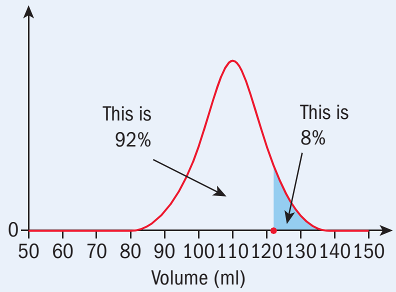
Sketch a diagram.
μ = mean = 110 g
σ = standard deviation = 8 g

Weight between 100 g and 130 g:
lower bound = 100
upper bound = 130

88.8% of the pears weigh between 100 g and 130 g.

From GDC:
normCdf(100, 130, 110, 8) = 0.888 = 88.8%

b



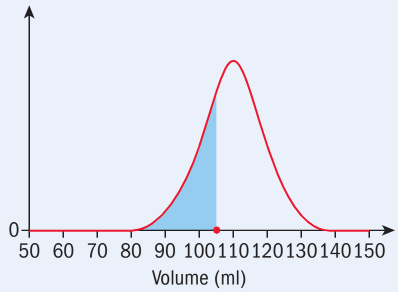
8% weighing more than m g is the same as saying that 92% weigh less than m g.

From GDC:
invNorm(0.92, 110, 8) = 121

m = 121 g

m = 121 g means that 8% of the pears weigh more than 121 g.

c



Weight less than 105 g:
lower bound = -9×10^{999}
upper bound = 105

P(weight less than 105 g) = 0.266
Expected number = 250 × 0.266 = 66.5
So, you would expect 66 or 67 pears to weigh less than 105 g.

From GDC:
normCdf(−9E999, 105, 110, 8) = 0.266
250 pears are weighed.

Exercise 5C

1 The mass of coffee grounds in Super-strength coffee bags is normally distributed with a mean of 5 g and a standard deviation of 0.1 g. It is known that 25% of the coffee bags weigh less than p grams. Find the value of p.

2 The heights of Dutch men are normally distributed with a mean of 181 cm and a standard deviation of 5 cm. It is known that 35% of Dutch men have a height less than h cm. Find the value of h.

3 The weight of kumquats is normally distributed with a mean of 20 g and a standard deviation of 0.8 g. It is known that 15% of the kumquats weigh more than k grams. Find the value of k.

4 The weight of cans of sweetcorn is normally distributed with a mean of 220 g and a standard deviation of 4 g. It is known that 30% of the cans weigh more than w grams. Find the value of w.

EXAM-STYLE QUESTIONS

5 The weights of cats are normally distributed with a mean of 4.23 kg and a standard deviation of 0.76 kg.

a Write down the weights of the cats that are within one standard deviation of the mean.

A vet weighs 180 cats.

b Find the number of these cats that would be expected to be within one standard deviation of the mean.

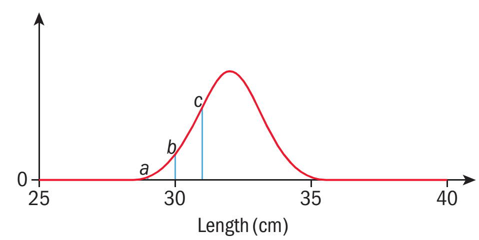
c Calculate the probability that a cat weighs less than 3.1 kg.

d Calculate the percentage of cats that weigh between 3 kg and 5.35 kg.

It is known that 5% of the cats weigh more than w kg.

e Find the value of w.

6 A manufacturer makes drumsticks with a mean length of 32 cm. The lengths are normally distributed with a standard deviation of 1 cm.



a Calculate the values of a, b and c shown on the graph.

b Find the probability that a drumstick has a length greater than 30.6 cm.

It is known that 80% of the drumsticks have a length less than d cm.

c Find the value of d.

One week 5000 drumsticks are tested.

d Calculate the expected number of drumsticks that have a length between 30.5 cm and 32.5 cm.

7 The average lifespan of a television set is normally distributed with a mean of 8000 hours and a standard deviation of 1800 hours.

a Find the probability that a television set will break down before 2000 hours.

b Find the probability that a television set lasts between 6000 and 12 000 hours.

c It is known that 12% of the television sets break down before t hours.
Find the value of t.

EXAM-STYLE QUESTIONS

8 The speed of cars on a motorway is normally distributed with a mean of $120\,\text{km}\,\text{h}^{-1}$ and a standard deviation of $10\,\text{km}\,\text{h}^{-1}$.

a Draw a normal distribution diagram to illustrate this information.

b Find the percentage of cars that are traveling at speeds of between $105\,\text{km}\,\text{h}^{-1}$ and $125\,\text{km}\,\text{h}^{-1}$.

It is known that 8% of the cars are traveling at a speed of less than $p\,\text{km}\,\text{h}^{-1}$.

c Find the value of p.

One day 800 cars are checked for their speed.

d Calculate the expected number of cars that will be traveling at speeds of between $96\,\text{km}\,\text{h}^{-1}$ and $134\,\text{km}\,\text{h}^{-1}$.

The speed limit is $130\,\text{km}\,\text{h}^{-1}$.

e Find the number of cars that are expected to be exceeding the speed limit.

9 The weights of bags of rice are normally distributed with a mean of 1003 g and a standard deviation of 2 g.

a Draw a normal distribution diagram to illustrate this information.

b Find the probability that a bag of rice weighs less than 999 g.

The manufacturer states that the bags of rice weigh 1 kg.

c Find the probability that a bag of rice is underweight.

400 bags of rice are weighed.

d Calculate the expected number of bags of rice that are underweight.

5% of the bags of rice weigh more than p g.

e Find the value of p.