4.

The normal distribution

- Standard scores
- Normal distribution
- Using a calculator
- In the old days – using a book of tables
- Notation
- Quantiles
- The normal distribution pdf
- Using the normal distribution to model data
- Can we use the normal distribution to model discrete data?
- Limitations of probability models for predicting real behaviour
- Miscellaneous exercise four

Note

Students taking both *Mathematics Methods and Mathematics Applications*, and who studied *Mathematics Applications Unit Two* earlier, will be familiar with much of the work of this chapter. However there are some ideas mentioned here that were not included in the treatment of the concept in *Mathematics Applications*, and some of the questions in this chapter were not in the corresponding chapter of *Mathematics Applications Unit Two*. Hence any students in this category are encouraged to work through this chapter anyway, as the repeat work will serve as useful revision, and the new work will extend their current understanding appropriately.

Situation

Test 1
27

Kym sits a Mathematics test and achieves a mark of 27.
In the next test she scores 30. Has she improved?

Test 2
30

Before answering this question we might first ask:
What was each test out of?

$\frac{27}{40}$

Suppose that test 1 was out of 40 and test 2 was out of 50.
Can we now decide whether she has improved?

$\frac{30}{50}$

Before answering we may want to know if the tests were of similar difficulty.
What was the mean mark in each test?

Mean
23

Suppose the mean in test 1 was 23 and in test 2 was 25.
Now can we judge whether her test 2 mark shows an improvement?

Mean
25

What if we also knew the standard deviation for each test as well?

St dev
5

Suppose the standard deviation in test 1 was 5 marks
and in test 2 was 10 marks.

St dev
10

Now can you suggest whether or not Kym's mark in test 2
was an improvement on her mark in test 1?

iStock.com/Thomas_EyeDesign

ISBN 9780170395205

Standard scores

In the *situation* on the previous page did you consider expressing Kym's test scores in terms of the number of standard deviations each was from the mean?

Expressing a score as a number of standard deviations above or below the mean is called **standardising** the score. We obtain the **standard score**.

$$\text{Standardised score} = \frac{\text{Raw score} - \text{mean}}{\text{Standard deviation}}$$

EXAMPLE 1

Jennifer scores 23, 35 and 17 in tests A, B and C respectively. If the mean and standard deviation in each of these tests are as given below express each of Jennifer's test scores as standardised scores.

Test A:	mean	30,	standard deviation	5
Test B:	mean	32,	standard deviation	6
Test C:	mean	15,	standard deviation	2.5

Solution

In Test A Jennifer's standardised score is $\frac{23-30}{5}$ i.e. -1.4.

In Test B Jennifer's standardised score is $\frac{35-32}{6}$ i.e. 0.5.

In Test C Jennifer's standardised score is $\frac{17-15}{2.5}$ i.e. 0.8.

Exercise 4A

1 Express each of the following as a standard score.

- **a** A score of 65 in a test that had a mean of 60 and a standard deviation of 5.
- **b** A score of 72 in a test that had a mean of 55 and a standard deviation of 10.
- **c** A score of 50 in a test that had a mean of 58 and a standard deviation of 4.
- **d** A score of 60 in a test that had a mean of 58 and a standard deviation of 4.
- **e** A score of 58 in a test that had a mean of 64 and a standard deviation of 8.

2 SuMin scores 30, 50, 7 and 26 in tests A, B, C and D respectively. If the mean and standard deviation in each of these tests are as given below, express each of SuMin's test scores as standardised scores.

Test A:	mean	20,	standard deviation	4
Test B:	mean	60,	standard deviation	10
Test C:	mean	6,	standard deviation	0.8
Test D:	mean	25,	standard deviation	5

ISBN 9780170395205

3 All of the first-year students on a particular technology course sat exams in the core subjects of Mathematics, Chemistry, Electronics and Computing. The exam results produced the following summary statistics:

Mathematics exam:	mean mark	60,	standard deviation	10.4
Chemistry exam:	mean mark	72,	standard deviation	7.2
Electronics exam:	mean mark	48,	standard deviation	14.6
Computing exam:	mean mark	63,	standard deviation	7.4

One student scored 56 in Mathematics, 74 in Chemistry, 39 in Electronics and 72 in Computing. Standardise each of these scores and rank the subjects for this student, listing them from best to worst on the basis of these standard scores.

4 All year ten students in a particular region sat exams in Mathematics, English, Science and Social Studies. The exam results in these subjects produced the following means and standard deviations.

Mathematics:	Mean	63,	Standard deviation	14
English:	Mean	64,	Standard deviation	10
Science:	Mean	72,	Standard deviation	8
Social Studies:	Mean	106,	Standard deviation	22

One student achieved the following scores:

76 in Mathematics,	75 in English,
78 in Science,	104 in Social Studies.

Rank the four subjects in order for this student, highest standardised score first.

5 Jill and her boyfriend Jack sit the same maths exam, along with the 156 other candidates studying the course for which the exam formed a part of the assessment.

- The exam was marked out of 120.
- The mean mark for the entire 158 students was 65.2 and the standard deviation of the marks was 8.8.
- Jill scored 74 out of 120 and Jack scored 63 out of 120.

Complete the three incomplete responses from Jill shown below in the following conversation between her and her mother:

Jill (arriving home from school): '*Hi Mum. How's your day been?*'

Jill's mum: '*Pretty good, dear. How was yours? Did you get any marks back from the exams you did?*'

Jill: '*Yeah, I got my maths mark.*'

Jill's mum: '*What did you get?*'

Jill, quoting her exam mark as a standard score replied: '*Well, I got _____.*'

Jill's mum: '*What! That sounds awful! What was the average?*'

Jill, again quoting standard scores: '*The mean was _____.*'

Jill's mum: '*What! What did Jack get?*'

Jill: '*Oh, he got _____.*'

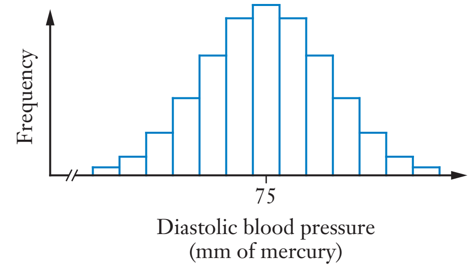
Jill's mum (who knew something about mathematics): '*Wait a minute. Are we talking standard scores here?*'

ISBN 9780170395205

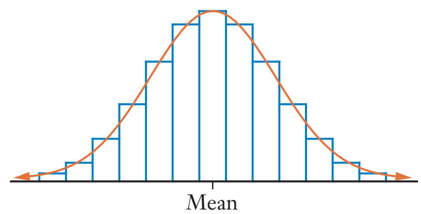
The standard normal curve

Normal distribution

Suppose the diastolic blood pressure of a large number of adults was measured and the mean value was found to be 75 mm of mercury (mm of mercury being the units blood pressure is measured in). The data collected, if presented as a histogram, could well have a shape similar to the diagram shown on the right, i.e. a symmetrical distribution with many values close to the mean and the number of values decreasing as we move further from the mean.



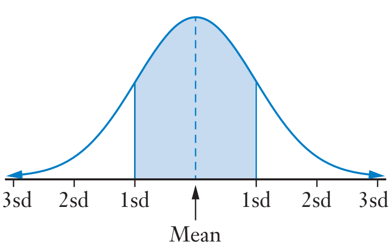
Fitting a smooth curve to the midpoints of the columns we obtain a '**bell-shaped curve**' as shown on the right.



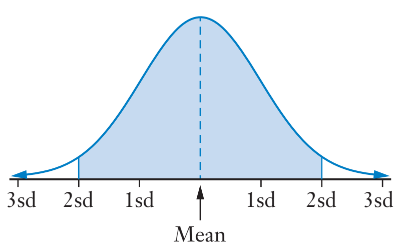
If we make many measurements of something that occurs naturally, for example, the heights of many adult females, the weights of many domestic cats, the foot lengths of many adult males, the histogram of the data often follows this sort of shape.

Data of this kind is said to be **normally distributed**. In **normal distributions** approximately two thirds of the population lie within one standard deviation of the mean, 95% would lie within two standard deviations of the mean and almost all would lie within three standard deviation of the mean.

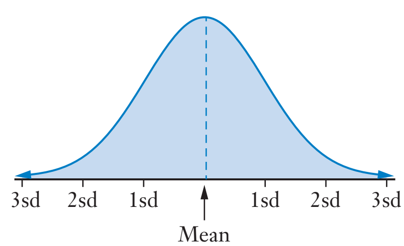
This is the 68%, 95%, 99.7% rule.



68.3% of area under curve shaded



95.4% of area under curve shaded



99.7% of area under curve shaded

In terms of probabilities, we could say that the probability of a randomly selected individual from a normally distributed population being within

- one standard deviation of the mean is 0.683,
- two standard deviations of the mean is 0.954,
- three standard deviations of the mean is 0.997.

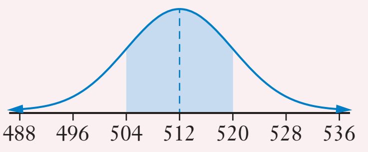
Note: The normal distribution is also referred to as the Gaussian distribution, after the German mathematician Carl Gauss.

ISBN 9780170395205

EXAMPLE 2

A box of breakfast cereal has 'contains 500 grams of breakfast cereal' printed on it. Suppose that in fact the weight of breakfast cereal contained in these boxes is normally distributed with a mean of 512 grams and a standard deviation of 8 grams. Determine the probability that a randomly chosen box of this cereal contains between 504 grams and 520 grams.

Solution



With a mean of 512 grams and a standard deviation of 8 grams:

504 grams is one standard deviation below the mean

and 520 grams is one standard deviation above the mean.

For normally distributed data the probability that a randomly chosen data point is within 1 standard deviation of the mean is, from the previous page, 0.683.

Thus the probability that a randomly chosen box of this cereal contains between 504 grams and 520 grams is 0.68.

The above example could be worked out using the '68' in the *68%, 95% 99.7% rule* because the question involved numbers of standard deviations that this rule relates to. What would we have done if instead the question had asked for the probability of a randomly chosen box of the cereal containing less than 500 grams? In this case 500 grams is 1.5 standard deviations below the mean, a situation not covered by the 68%, 95% 99.7% rule. In this case we can use the ability of various calculators to determine such probabilities, as shown by the next example (also based on the breakfast cereal situation of the above example).

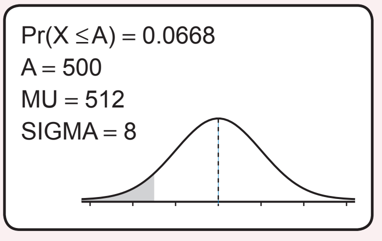
EXAMPLE 3

A box of breakfast cereal has 'contains 500 grams of breakfast cereal' printed on it. Suppose that in fact the weight of breakfast cereal contained in these boxes is normally distributed with a mean of 512 grams and a standard deviation of 8 grams.

a Determine the probability that a randomly chosen box of this cereal contains less than 500 grams.

b In a random sample of 100 boxes of this cereal approximately how many boxes should we expect to contain less than 500 g?

Solution



a For a randomly distributed set of values, with mean 512 and standard deviation 8, we require P(Randomly chosen value < 500).

Many calculators can display such information for normally distributed data.

The required probability is 0.0668.

The probability that a randomly chosen box of this cereal contains less than 500 grams is 0.0668.

b In any batch of boxes of this cereal we should expect that the proportion of them that contain less than 500 grams is about 0.07. Thus in a random sample of 100 boxes of this cereal we would expect approximately 7 boxes to contain less than 500 g.

ISBN 9780170395205

Using a calculator

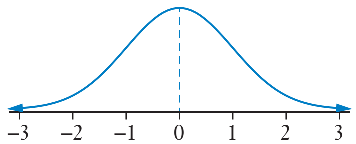
The various calculators have different capabilities and routines with regard to displaying probabilities for normally distributed sets of data.

You will gain familiarity with the ability of *your* calculator in this regard in the next exercise.

In the old days – using a book of tables

Prior to the ready availability of calculators with built-in statistical routines for determining probabilities associated with normal distributions, these probabilities were determined using books of statistical tables.

These books give probabilities for just one normal distribution, the **standard normal distribution**. For this the random variable has a mean of 0 and a standard deviation of 1, as shown on the right.



Normal distributions having means and standard deviation not equal to these standard values needed to be standardised. We encountered this idea of standardising data by expressing it as a number of standard deviations above or below the mean at the beginning of this chapter. Calling the original score an 'x score' and the standardised score a 'z score' we have:

$$z \text{ score} = \frac{x \text{ score} - \text{mean of } x \text{ scores}}{\text{standard deviation of } x \text{ scores}}$$

Thus before the ready availability of sophisticated calculators, to answer the previous example which required us to determine the probability that from a normally distributed set of data, X, with mean 512 and standard deviation 8, a randomly selected item would have a value less than 500 we would have changed the 500 to a standard score:

$$\begin{aligned} \text{standard score} &= \frac{500 - 512}{8} \\ &= -1.5 \end{aligned}$$

(i.e. a score of 500 is 1.5 standard deviations below the mean)

and then used the table of probabilities for the standard normal distribution to determine the required probability.

$$\begin{aligned} P(X < 500) &= P(Z < 1.5) \\ &= 0.0668 \end{aligned}$$

z	0.00	0.01	0.02	0.03
−1.9	0.0287	0.0281	0.0274	0.0268
−1.8	0.0359	0.0351	0.0344	0.0336
−1.7	0.0446	0.0436	0.0427	0.0418
−1.6	0.0548	0.0537	0.0526	0.0516
−1.5	0.0668	0.0655	0.0643	0.0630

Thus, as before, the probability that a randomly chosen box of the cereal contains less than 500 grams is 0.0668.

ISBN 9780170395205

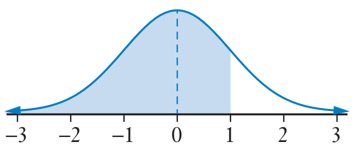
Exercise 4B

The questions of this exercise refer to data sets involving normally distributed scores, X.

Using your calculator make sure that you can obtain each of the probabilities given in questions **1** to **8** below (correct to 4 decimal places), and each value of k in questions **9** to **17**.

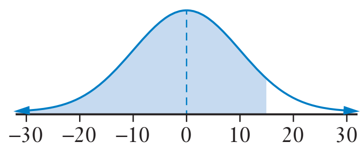
Get to know the capabilities of your calculator with regard to normal probability distributions.

1 mean = 0
standard deviation = 1



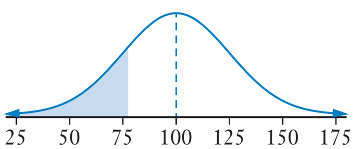
$P(X < 1) = 0.8413$
Can you also get 0.84 using the 68%, 95%, 99.7% rule?

2 mean = 0
standard deviation = 10



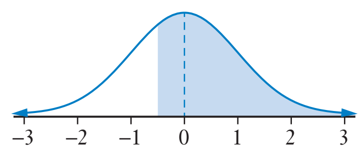
$P(X < 15) = 0.9332$

3 mean = 100
standard deviation = 25



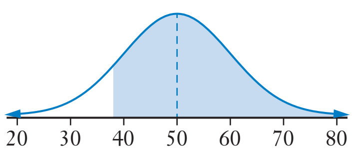
$P(X < 78) = 0.1894$

4 mean = 0
standard deviation = 1



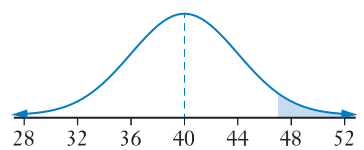
$P(X > -0.5) = 0.6915$

5 mean = 50
standard deviation = 10



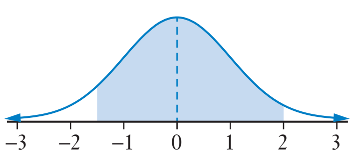
$P(X > 38) = 0.8849$

6 mean = 40
standard deviation = 4



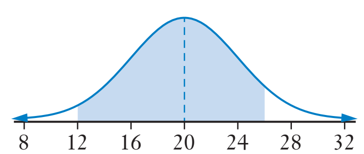
$P(X > 47) = 0.0401$

7 mean = 0
standard deviation = 1



$P(-1.5 < X < 2) = 0.9104$

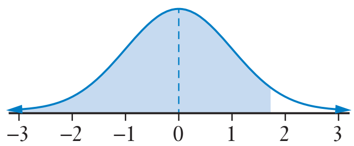
8 mean = 20
standard deviation = 4



$P(12 < X < 26) = 0.9104$

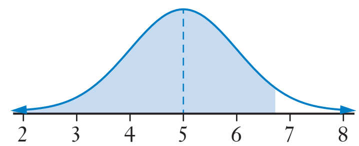
ISBN 9780170395205

9 mean = 0
standard deviation = 1



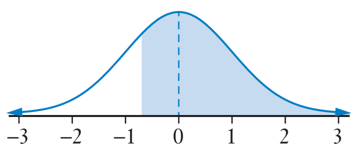
$P(X < k) = 0.9573$
$\therefore k = 1.72$ (2 decimal places)

10 mean = 5
standard deviation = 1



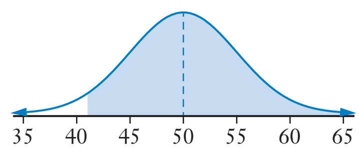
$P(X < k) = 0.9671$
$\therefore k = 6.84$ (2 decimal places)

11 mean = 0
standard deviation = 1



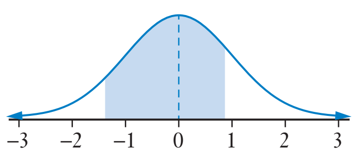
$P(X > k) = 0.7517$
$\therefore k = -0.68$ (2 decimal places)

12 mean = 50
standard deviation = 5



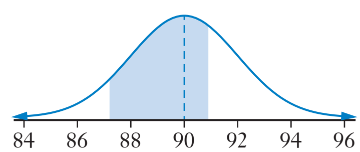
$P(X > k) = 0.9656$
$\therefore k = 40.9$ (1 decimal place)

13 mean = 0
standard deviation = 1



$P(-1.4 < X < k) = 0.7215$
$\therefore k = 0.85$ (2 decimal places)

14 mean = 90
standard deviation = 2



$P(87.2 < X < k) = 0.5964$
$\therefore k = 90.92$
(2 decimal places)

15
mean = 40
standard deviation = 15
$P(X < k) = 0.9850$
$\therefore k = 72.55$ (2 decimal places)

16
mean = 10
standard deviation = 0.5
$P(X > k) = 0.0721$
$\therefore k = 10.73$
(2 decimal places)

17
mean = 0.1
standard deviation = 0.01
$P(0.08 < X < k) = 0.3036$
$\therefore k = 0.0955$ (4 decimal places)

Notation

If we use X to represent the possible values of a normally distributed set of measurements having a mean μ and standard deviation σ (and hence variance σ^2) this is sometimes written:

$$X \sim N(\mu, \sigma^2).$$

With μ (mu) pronounced ‘myew’, this is read as:

X is normally distributed with mean myew and standard deviation sigma.

ISBN 9780170395205

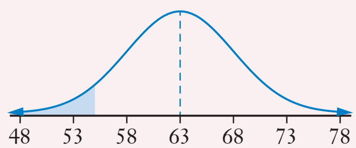
EXAMPLE 4

If $X \sim N(63, 25)$ determine $P(X < 55)$.

Solution

X is normally distributed with a mean of 63 and a standard deviation of 5.

Using a calculator:
$P(X < 55) = 0.0548$



Using a tables book:
$P(X < 55) = P(Z < -1.6)$
$= 0.0548$
(Shown for interest only.)

Quantiles

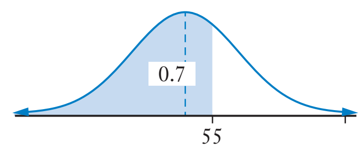
Normal distribution – Worded problems 1

Normal distribution – Worded problems 2

Quantiles are the values which a particular proportion of the distribution falls below.

Thus if 0.7 (70%) of the distribution is below 55 then 55 is the 0.7 quantile.

Alternatively we can refer to 55 as being the 70th **percentile**.



Note:
- We are already accustomed to referring to the 0.25 quantile as the first, or lower, **quartile** and the 0.75 quantile as the third, or upper, quartile.
- If the quartiles divide a distribution in to four equal parts and the percentiles divide the distribution into 100 equal parts, what might deciles and quintiles do?

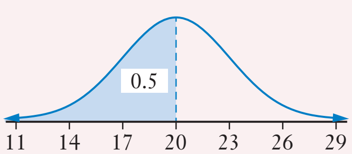
EXAMPLE 5

If $X \sim N(20, 3^2)$ determine

a the 0.5 quantile,
b the 0.82 quantile,
c the 24th percentile,
d the 62nd percentile.

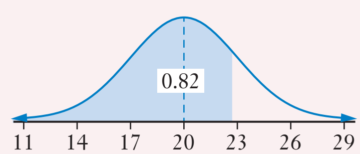
Solution

a



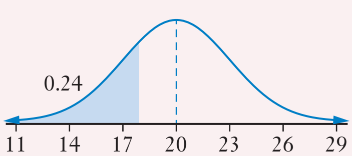
By inspection:
The 0.5 quantile is 20.

b



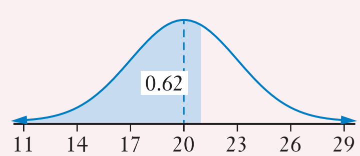
Using a calculator:
The 0.82 quantile is 22.7.

c



Using a calculator:
The 24th percentile is 17.9.

d



Using a calculator:
The 62nd percentile is 20.9.

ISBN 9780170395205

EXAMPLE 6

Eight thousand, two hundred and forty students were given an IQ test. The scores were normally distributed with a mean of 100 and a standard deviation of 16.

a Determine how many of the students, to the nearest ten, achieved a score in excess of 128.

b What were the minimum and maximum scores of the middle 60% of students on this test?

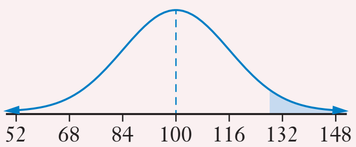
Solution

a Let X be the scores obtained in the test.

Thus $X \sim N(100, 16^2)$.

We require $P(X > 128)$.

Using a calculator, $P(X > 128) = 0.0401$.



Number scoring more than 128:

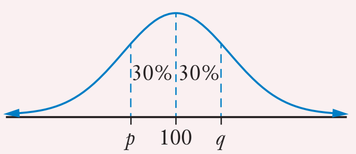
$$0.0401 \times 8240 \approx 330$$

Approximately 330 students achieved a score in excess of 128.

b If p is the lowest score achieved by the middle 60%

then $P(X < p) = 0.2$ i.e. $p = 86.53$

and if q is the highest score achieved by the middle 60% then $P(X < q) = 0.8$ i.e. $q = 113.47$



(Some calculators can determine p and q more directly for this symmetrical situation.)

The lowest and highest scores achieved by the middle 60% are 86.5 and 113.5 respectively (to the nearest half-mark).

EXAMPLE 7

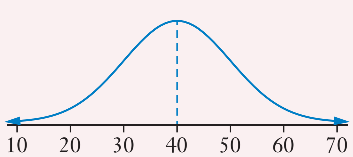
If $X \sim N(40, 10^2)$ determine each of the following probabilities using the 68%, 95%, 99.7% rule, and *not* the statistical capability of your calculator.

a $P(30 < X < 50)$ **b** $P(20 < X < 60)$ **c** $P(40 < X < 60)$ **d** $P(X \leq 50)$

Solution

a 30 is one standard deviation below the mean and 50 is one standard deviation above the mean.

Thus $P(30 < X < 50) = 0.68$



b $P(20 < X < 60) = 0.95$

c $P(40 < X < 60) = \dfrac{0.95}{2}$

$= 0.48$ (correct to 2 decimal places)

d $P(X \leq 50) = 0.5 + \dfrac{0.68}{2}$

$= 0.84$

As the previous chapter explained, we make no distinction between $P(X \leq 50)$ and $P(X < 50)$. Including the line or not makes no difference to the area of the region.

ISBN 9780170395205

EXAMPLE 8

Let us suppose that the time from Simon getting out of bed until his arrival at school is normally distributed with a mean of 55 minutes and a standard deviation of 5 minutes. Simon's arrival at school is classified as being late if it occurs after 9:10 a.m.

a One day Simon gets out of bed at 8:08 a.m. What is the probability of him arriving late?

b For a period of time Simon always gets out of bed at the same time but finds that he arrives late approximately 85% of the time! What time is he getting out of bed (to the nearest minute)?

Solution

a Let T minutes be the time from getting out of bed until arrival at school.

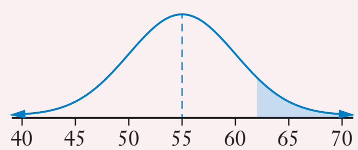
Thus $T \sim N(55, 5^2)$.

Simon has 62 minutes to get to school before he is late.

We require: $P(T > 62)$

Calculator gives: $P(T > 62) = 0.0808$.

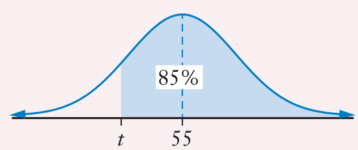
If Simon gets out of bed at 8:08 a.m. the probability of him arriving late is 0.0808.



b The time that Simon is allowing himself to get to school is causing him to be late approximately 85% of the time.

We require t for which $P(T > t) = 0.85$.

Calculator gives: $t \approx 49.8$

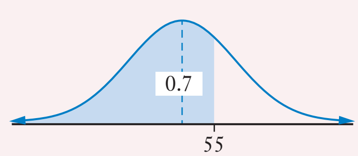


Thus Simon is allowing approximately 50 minutes to get to school and for 85% of the days the journey takes longer than this, causing him to be late 85% of the time.

Simon is getting out of bed at 8:20 a.m.

EXAMPLE 9

The continuous random variable, X, is normally distributed with $P(X < 55) = 0.7$.



a How many standard deviations from the mean is a score of 55?

b If the standard deviation of X is 4 find the mean of the distribution, giving your answer correct to one decimal place.

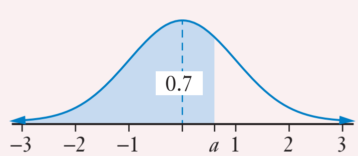
Solution

a The standard normal distribution shows standard deviations from the mean (see diagram).

Thus $Z \sim N(0, 1)$ and $P(Z \le a) = 0.7$

From calculator $a = 0.5244$

55 is 0.524 standard deviations from the mean.



b $\text{Mean} + 0.524 \times 4 = 55$

$\therefore \text{Mean} = 52.9$ correct to one decimal place.

ISBN 9780170395205

EXAMPLE 10

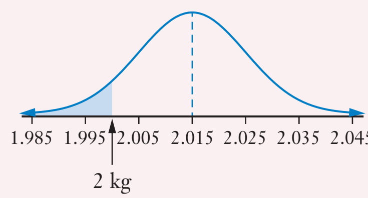
Pre-bagged packs of bananas are marked as 'contains approximately 2 kg'. Let us suppose that in fact that the weight of the contents of such bags are normally distributed with mean 2.015 kg and standard deviation 0.01 kg.

a What is the probability that the contents of a randomly chosen bag of these bananas has a weight of less that 2 kg?

b If five such bags are randomly selected what is the probability that three or more will have contents weighing less than 2 kg?

Solution

a Let the weight of the bags be represented by the continuous random variable X.



Thus $X \sim N(2.015, 0.01^2)$.

Calculator gives $P(X < 2) = 0.066\,807\,2$

The probability that the contents of a randomly chosen bag of the bananas having a weight of less than 2 kg is 0.0668 (correct to 4 decimal places).

b We now consider a binomial distribution Y with $Y \sim \text{Bin}(5, 0.066\,81)$.

We require $P(Y \geq 3)$.

By calculator, as shown below:

```
binomialCdf(5,0.06681,3,5)
                  0.002691
```

Or as shown in the working below,

$$P(Y \geq 3) = {}^5C_3\,(0.066\,81)^3\,(1 - 0.066\,81)^2 + {}^5C_4\,(0.066\,81)^4\,(1 - 0.066\,81)^1 + (0.066\,81)^5$$
$$= 0.0027 \text{ (correct to 4 decimal places).}$$

Thus if five of the bags are randomly selected the probability that three or more will have contents weighing less that 2 kg is 0.0027.

ISBN 9780170395205

The normal distribution pdf

Though not something you necessarily need to know, but included here for interest, is the fact that the probability density function, $f(x)$, for a normal distribution with a mean of μ and standard deviation σ, i.e. $X \sim N(\mu, \sigma^2)$, is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-0.5\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{for } -\infty < x < \infty.$$

Thus for the standard normal distribution, $X \sim N(0, 1^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-0.5x^2} \qquad \text{for } -\infty < x < \infty.$$

We would then expect $\displaystyle\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-0.5x^2}\,dx = 1.$

Confirm this result on your calculator but note that $f(x)$ is not readily integrated by any methods we have encountered. Instead, and beyond the scope of this unit, your calculator may use a numerical method, and may give an answer very close to 1, due to the numerical approximation involved in the method.

Taking, as an example, a normally distributed random variable X, with mean 50 and a standard deviation of 10, i.e. $X \sim N(50, 10^2)$, we can (with the assistance of a calculator) check that the formula given at the top of this page does give an answer consistent with our understanding that approximately two-thirds of the population lie within one standard deviation of the mean:

$$\begin{aligned} P(40 \le X < 60) &= \int_{40}^{60} \frac{1}{10\sqrt{2\pi}} e^{-0.5\left(\frac{x-50}{10}\right)^2} dx \\ &\approx 0.683. \end{aligned}$$

Fortunately, and as we have already seen, we do not have to formulate this definite integral each time, instead obtaining the answer from the ability of many calculators to output such values for normal distributions.

```
normCDf(40,60,10,50)
                0.6826894921
```

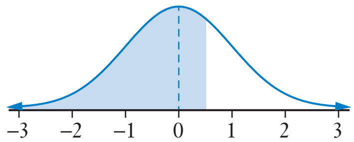
ISBN 9780170395205

Exercise 4C

Questions **1** to **8** of this exercise refer to data sets that involve normally distributed scores, X.

1

mean = 0
standard deviation = 1

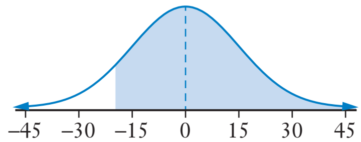


$P(X < 0.5) = k$

Find k correct to 4 decimal places.

2

mean = 0
standard deviation = 15

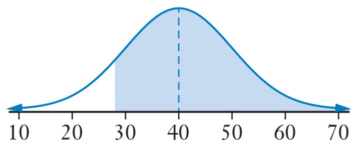


$P(X > -20) = k$

Find k correct to 4 decimal places.

3

mean = 40
standard deviation = 10

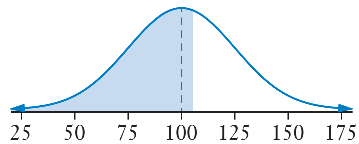


$P(X > 28) = k$

Find k rounded to 4 decimal places.

4

mean = 100
standard deviation = 25

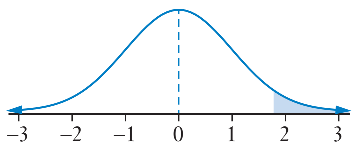


$P(X < 105) = k$

Find k rounded to 4 decimal places.

5

mean = 0
standard deviation = 1

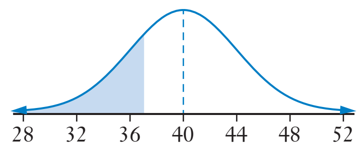


$P(X > k) = 0.0418$

Find k rounded to 2 decimal places.

6

mean = 40
standard deviation = 4

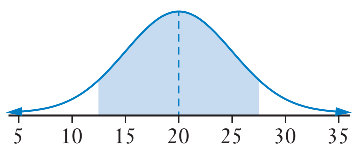


$P(X < k) = 0.2776$

Find k rounded to 2 decimal places.

7

mean = 20
standard deviation = 5

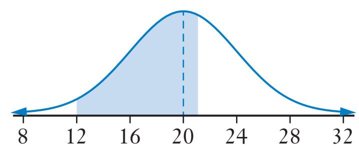


$P(20 - k < X < 20 + k) = 0.8684$

Find k rounded to 2 decimal places.

8

mean = 20
standard deviation = 4



$P(12 < X < k) = 0.6$

Find k rounded to 2 decimal places.

ISBN 9780170395205

9 The random variable, X, is normally distributed with a mean of 12 and a standard deviation of 2, i.e. $X \sim N(12, 2^2)$. Determine $P(X \geq 13.5)$.

10 The random variable, X, is normally distributed with a mean of 240 and a variance of 400, i.e. $X \sim N(240, 20^2)$. Determine $P(218 < X < 255)$.

11 $X \sim N(62, 64)$, i.e. X, is normally distributed with a mean of 62 and a standard deviation of 8. Given that $P(X > k) = 0.8238$ determine k.

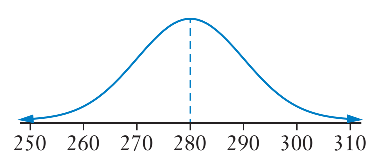
12 If $X \sim N(0, 1)$ determine
a the 0.72 quantile,
b the 0.26 quantile,
c the 89th percentile,
d the 23rd percentile.

13 If $X \sim N(20, 3^2)$ determine
a the 0.44 quantile,
b the 0.74 quantile,
c the 33rd percentile,
d the 85th percentile.

14 Using the 68%, 95%, 99.7% rule, and *not* the statistical capability of your calculator, determine the following probabilities.

a $P(-1 < X < 1), X \sim N(0, 1^2)$.
b $P(-2 < X < 2), X \sim N(0, 1^2)$.
c $P(-3 < X < 3), X \sim N(0, 1^2)$.
d $P(8 < X < 32), X \sim N(20, 6^2)$.
e $P(4 < X < 16), X \sim N(10, 2^2)$.
f $P(0 < X < 1), X \sim N(0, 1^2)$.
g $P(X < 1), X \sim N(0, 1^2)$.
h $P(X > 1), X \sim N(0, 1^2)$.
i $P(X < 5), X \sim N(0, 5^2)$.
j $P(X > 70), X \sim N(60, 10^2)$.

15 Let us suppose that the duration of pregnancy, for a naturally delivered human baby, is a normally distributed variable with a mean of 280 days and a standard deviation of 10 days.



Using the 68%, 95%, 99.7% rule, and *not* the statistical capability of your calculator, determine estimates for the following.

a The percentage of human pregnancies, for naturally delivered babies, that are between 250 days and 310 days.
b The percentage of human pregnancies, for naturally delivered babies, that exceed 290 days.
c The percentage of human pregnancies, for naturally delivered babies, that are between 260 days and 270 days.

ISBN 9780170395205

16 A machine produces components whose weights are normally distributed with a mean of 500 g and standard deviation of 5 g.

a According to the 68%, 95%, 99.7% rule, what percentage of the components will have a weight of less than 495 g?

b According to the 68%, 95%, 99.7% rule, what percentage of the components will have a weight of less than 490 g?

17 A box of breakfast cereal has 'contains 300 grams of breakfast cereal' printed on it. Suppose that in fact the weight of breakfast cereal contained in these boxes is normally distributed with a mean of 310 grams and a standard deviation of 4 grams. Determine the probability that a randomly chosen box of this cereal contains

a more than 312 grams of breakfast cereal,

b less than 300 grams of breakfast cereal.

18 The lengths of adult male lizards of a particular species are thought to be normally distributed with a mean of 17.5 cm and a standard deviation of 2.5 cm.

Determine the probability that a randomly chosen adult male lizard of this species will have a length

a less than 17.5 cm,

b between 15 cm and 17.5 cm.

19 The scaled scores in a national mathematics test are normally distributed with a mean of 69 and a standard deviation of 12.

What is the probability that a randomly selected candidate who sat this test has a scaled score of

a more than 75?

b between 66 and 75?

c less than 45?

20 The heights of fully grown plants of a certain species are normally distributed with a mean of 30 cm and a standard deviation of 4 cm. If 100 fully grown plants of this species are randomly selected, approximately how many would you expect to be:

a taller than 35 cm?

b shorter than 25 cm?

c between 25 cm and 30 cm in height?

21 Let us suppose that 44 mg is 110% of the recommended daily intake of a particular vitamin and that a 110 mL container of fruit juice contains approximately 44 mg of this vitamin. If in fact the weight of the vitamin in the 110 mL containers of the fruit juice is normally distributed with mean 44 mg and standard deviation 2.5 mg, determine the probability that a randomly chosen 110 mL container of this fruit juice contains less than the recommended daily intake of the vitamin.

ISBN 9780170395205

22 Five thousand, five hundred and forty-two students sat a particular leaving exam that was marked out of 120. The scores obtained could be well modelled by a normal distribution with a mean of 62 and a standard deviation of 12.5.

a Distinction certificates were awarded to students who gained a mark of 80 or more. How many students gained distinction certificates?

b A mark of less than 40 was regarded as a fail. How many of the students failed?

c What were the minimum and maximum scores of the middle 40% of students on this test?

23 Let us suppose that the heights of the adults of a particular country are normally distributed with a mean of 1.75 m and a standard deviation of 10 cm. A car manufacturer wishes to design a new car with the space allowed for the driver, and the 'travel' on the drivers seat, suitable for every adult in the population except the tallest 5% of the adult population and the shortest 5% of the adult population. What is the height of the shortest driver and the tallest driver that the manufacturer is attempting to allow for? (Answer to nearest half-centimetre.)

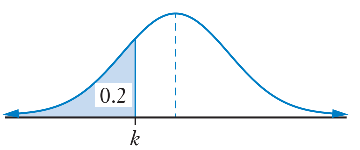
24 The marks achieved in a particular exam are normally distributed with a mean of 64 and a standard deviation of 12.

Grades are to be awarded as follows:

Top	12%	of candidates:	Grade A
Next	25%	of candidates:	Grade B
Next	40%	of candidates:	Grade C
Next	15%	of candidates:	Grade D
Remainder		of candidates:	Grade F

Determine the marks that form the A/B, B/C, C/D, and D/F grade boundaries, giving your answers correct to the nearest whole number.

25 The continuous random variable, X, is normally distributed with $P(X < k) = 0.2$.



a How many standard deviations from the mean is k?

b If $k = 40$, and the standard deviation of X is 5, find the mean of the distribution, giving your answer correct to one decimal place.

26 Let us suppose that the time, in minutes, from Monica leaving home until she arrives at work is a normally distributed random variable with a mean of 45 and a standard deviation of 5. Monica's arrival at work is classified as late if it occurs after 8:30 a.m.

a One day Monica leaves home at 7:40 a.m. What is the probability of her arriving late?

b For a period of time Monica leaves home at the same time each day. During this period she finds that she arrives late approximately 8% of the time. What time is she leaving home (to the nearest minute)?

c What is the latest time (involving whole minutes) that Monica should leave home each day if she wishes to cut her late arrivals to less than 1%?

ISBN 9780170395205

27 The annual rainfall in an area in the south west of Western Australia is normally distributed with a mean of 1200 mm and a standard deviation of 200 mm.

According to this model, and assuming the situation does not change, in every one hundred years how many years would you expect the annual rainfall to be

a less than 800 mm?

b more than 1500 mm?

c between 800 mm and 1500 mm?

d Given that a year has an annual rainfall of more than 1300 mm, what is the probability that the rainfall for the year is less than 1500 mm?

28 The weight of each apple harvested from a particular orchard determines where the apple will be sent:

If	weight of apple	$\geq$	250 g	send to premium outlet
	150g < weight of apple	$<$	250 g	send to normal market
	weight of apple	$\leq$	150 g	send for juicing.

The weights of the apples are normally distributed with mean 180 g and standard deviation 40 g.

a In a random sample of 1000 apples how many would you expect to go to the premium outlet?

b Given that an apple does not go to the premium outlet, what is the probability that it is sent for juicing?

29 Bags of tomatoes are marked as
Contains approximately 1 kg tomatoes.

An analysis indicated that the weights of the tomatoes in the bags were approximately normally distributed with a mean of 1018 grams and standard deviation 10 grams.

Based on this normal distribution, what is the probability that a randomly chosen bag will contain tomatoes weighing

a less than 1 kg?

b at least 25 grams over 1 kg?

c If ten bags were randomly chosen, what is the probability that at least one would weigh less than 1 kg? (Give this answer correct to two decimal places.)

Shutterstock.com/iLight photo

30 Let us suppose that a particular IQ test gives results that are normally distributed with a mean score of 100 and a standard deviation of 15.

a If a randomly chosen person has a score on this test that is greater than 125, what is the probability their score is greater than 135?

b If five people were randomly selected, what is the probability that at least three would have a score greater than 120?

ISBN 9780170395205

31 Part of a breakfast cereal packing process involves a machine sending x grams of the cereal into each packet. The value of x is programmed into the machine and the packets are then filled with the weight of cereal in each being normally distributed with a mean of x grams and a standard deviation of 1.8 grams.

a The machine is used to fill packets that will be labelled as containing 500 grams. What should be the value of x, rounded up to the next gram, if the company wants no more than 0.5% of the packets to be underweight?

b The machine is used to fill packets that will be labelled as containing 250 grams. What should be the value of x, rounded up to the next gram, if the company wants no more than 0.5% of the packets to be underweight?

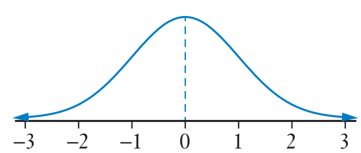
32 A machine produces components whose weights are normally distributed with a mean of 500 g and standard deviation of 5 g.

a What percentage of these components will have a weight of less than 490 g?

A new machine is being developed to produce these components more uniformly. The intention is for this new machine to produce components whose weights are normally distributed with a mean of 500 g and just 1.5% having a weight of less than 490 g.

b What does the standard deviation of the weights of components made by this new machine need to be?

Using the normal distribution to model data



The normal distribution is an extremely useful distribution and is used to model many naturally occurring random variables, for example, the blood pressures of the adult male population or adult female population of a country.

However, given a set of collected data, how would we know if it was appropriate to model the distribution of the data as a normal distribution?

One way to check the appropriateness would be to view the histogram of the distribution to see if it has the characteristic bell shaped curve.

We can also check to see if the data set gives proportions of data values lying within particular ranges similar to the proportions we would expect from a normally distributed random variable.

For example, suppose we have a distribution of scores with mean 17.2 and standard deviation 2.1. If this distribution were normally distributed we would expect approximately 68% of the scores to lie within one standard deviation of the mean.

$$\text{i.e. for } X \sim N(17.2, 2.1^2),\ P(15.1 < X < 19.3) \approx 0.68$$

If, for our data set, the proportion was not close to 68% we would question the wisdom of modelling the distribution as a normal distribution.

Can we use the normal distribution to model discrete data?

Whilst the normal distribution is for continuous data it can be used to model discrete data if we make a 'correction, or adjustment, for continuity'.

To determine $P(8 \leq X \leq 10)$ for a discrete distribution of integers X we would determine

$$P(7.5 < Y < 10.5)$$

where Y is a suitably chosen continuous variable.

Similarly	$P(X < 50)$	would become	$P(Y < 49.5)$	on a continuous model,
	$P(X \leq 50)$	would become	$P(Y < 50.5)$	on a continuous model,
and	$P(8 < X < 10)$	would become	$P(8.5 < Y < 9.5)$	on a continuous model.

Limitations of probability models for predicting real behaviour

If we model a particular probabilistic situation as normal, binomial, uniform, etc. we must remember that the model is still only a model. The real situation may not fit the model exactly.

Our data may not be truly representative of the real situation.

The model may fit the situation generally but be 'a bad fit' in specific circumstances.

Influences and events that we may assume to be random may not be totally random.

Etc.

Thus if we record data and then model the distribution of this data as being of a particular type and with particular characteristics, any predictions we make based on the model need to be viewed with some caution. For example, suppose we were to collect data from adult female Australians and suppose this data suggested it appropriate to model the blood pressures of adult Australian females as being normally distributed, with a particular mean and standard deviation This may not be an appropriate model to use if we were applying it to a group of super fit adult females or to a group of elderly adult females or to adult females from other countries or if our data was based on a small sample, or a biased sample or … .

Exercise 4D

Explain why each of the situations in numbers **1** to **5** would cause you to question the wisdom of choosing the normal distribution as a model for the data.

1 A data set has 80% of its data points within one standard deviation of the mean.

2 A data set involved 360 measurements of continuous data with mean 4.37 and standard deviation 2.52.
Fifty-two of the 360 measurements were greater than 8.

3 A data set involved 826 measurements of a continuous variable with mean 8.9 and standard deviation 3.1.
Two hundred and fifty-seven of the measurements were within one standard deviation of the mean.

ISBN 9780170395205

4 A data set involved 409 measurements of a continuous variable.

One hundred and five of the measurements were between the mean and one standard deviation below the mean.

One hundred and eighty of the measurements were between the mean and one standard deviation above the mean.

5 A data set involved 180 measurements with mean 105.2 and standard deviation 31.4.

Ten of the measurements were at least two standard deviations above the mean.

None of the measurements were more than two standard deviations below the mean.

6 A scientist analyses the 520 measurements made of a particular continuous variable, X. The scientist finds that the measurements have a mean of 26.4, a standard deviation of 3.7 and can be grouped as follows:

15.3 – 19.0	19.0 – 22.7	22.7 – 26.4	26.4 – 30.1	30.1 – 33.8	33.8 – 37.5
8 measurements	72 measurements	181 measurements	173 measurements	74 measurements	12 measurements

The scientist decides to use $X \sim N(26.4, 3.7^2)$ to predict probabilities if the number of measurements in the range

mean $\rightarrow$ mean + 1 st. devn.	mean – 1 st. devn. $\rightarrow$ mean
mean $\rightarrow$ mean + 2 st. devns	mean – 2 st. devns $\rightarrow$ mean
mean $\rightarrow$ mean + 3 st. devns	mean – 3 st. devns $\rightarrow$ mean

are all within 3% of the numbers predicted by $X \sim N(26.4, 3.7^2)$ (This is not any standard test, just one she decides to apply.)

a Will she use $X \sim N(26.4, 3.7^2)$?

b Does her 3% rule ensure that, if it is met, any predicted probabilities will be sufficiently reliable?

7 A set of data has the following summary statistics:

mean = 44.7 median = 44.9 standard deviation = 26.4

lower quartile = 19.3 upper quartile = 69.2

State, with reasoning, whether these figures suggest the data is normally distributed or not.

Shutterstock.com/Delpixel

8 For large values of n the binomial probability distribution $X \sim \text{Bin}(n, p)$ can be well modelled by the normal distribution $Y \sim \text{N}(np, np(1-p))$.

I.e., for large n, binomial probabilities involving n trials, with the probability of success at each trial being p, can be well modelled by a normal distribution with mean np and standard deviation $\sqrt{np(1-p)}$, with appropriate adjustments for continuity being made. (See page 96 for an explanation of adjustment for continuity.)

a Use you calculator to determine $P(X \leq 65)$ for $X \sim \text{Bin}(100, 0.6)$ and $P(Y \leq 65.5)$ for $Y \sim \text{N}(60, 24)$ giving each answer correct to four decimal places.

b Use your calculator to determine $P(X = 100)$ for $X \sim \text{Bin}(200, 0.5)$ and $P(99.5 \leq Y \leq 100.5)$ for the appropriate normal distribution model, giving each answer correct to four decimal places.

c Use your calculator to determine $P(20 < X \leq 25)$ for $X \sim \text{Bin}(50, 0.6)$ and the equivalent probability on the appropriate normal distribution model, giving each answer correct to four decimal places.

Miscellaneous exercise four

This miscellaneous exercise may include questions involving the work of this chapter, the work of any previous chapters, and the ideas mentioned in the Preliminary work section at the beginning of the book.

Differentiate each of numbers **1** to **6** with respect to x. For some it may be advisable to use the laws of logarithms before differentiating.

1 $y = \ln(10x)$

2 $y = 10\ln x$

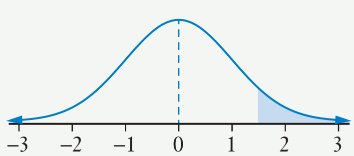
3 $y = \dfrac{x}{\ln x}$

4 $y = \log_e\left[(x^2+1)^3\right]$

5 $y = \ln\left[\dfrac{(x-1)^3}{x+1}\right]$

6 $y = \log_5 x$

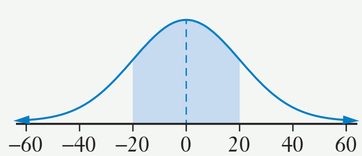
7 Normal distribution,
Mean = 0,
Standard deviation = 1.



$P(X > 1.5) = k$

Find k correct to 4 decimal places.

8 Normal distribution,
Mean = 0,
Standard deviation = 20.

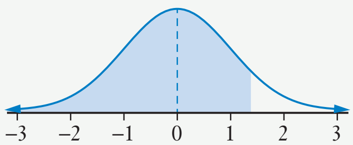


$P(-20 < X < 20) = k$

Find k correct to 4 decimal places.

ISBN 9780170395205

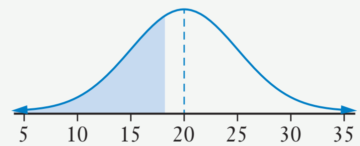
9 Normal distribution,
Mean = 0,
Standard deviation = 1.



$P(X < k) = 0.9066$

Find k rounded to 2 decimal places.

10 Normal distribution,
Mean = 20,
Standard deviation = 5.



$P(X < k) = 0.3632$

Find k rounded to 2 decimal places.

Find the exact gradient of each of the following at the given point on the curve.

11 $y = 3\ln x$ at $(e, 3)$.

12 $y = x\ln x$ at (e, e).

13 The random variable, X, is normally distributed with a mean of 50 and a standard deviation of 3, i.e. $X \sim N(50, 3^2)$. Determine $P(X \geq 58)$.

14 If $X \sim N(0, 1)$ determine, to three decimal places:

a the 0.42 quantile,

b the 0.13 quantile,

c the 63rd percentile,

d the 41st percentile.

15 Let us suppose that the daily rainfall in a region is normally distributed with a mean of 11.2 mm and a standard deviation of 3.1 mm.

In a year of 365 days, how many days would you expect this region to have a rainfall that is

a less than 6 mm?

b more than 10 mm?

c between 10 mm and 15 mm?

Shutterstock.com/MK photograp55

16 A continuous random variable, X, has pdf:

$$f(x) = \begin{cases} ax^2 + k & \text{for } 0 \leq x \leq 2 \\ 0 & \text{elsewhere.} \end{cases}$$

If $P(X \leq 1) = 0.2$, determine a and k.

Hence find $E(X)$, the expected value of X, and $Var(X)$, the variance of X.

ISBN 9780170395205

17 A continuous random variable, X, has pdf:

$$f(x) = \begin{cases} k\sin x & \text{for} \quad 0 \le x \le \pi \\ 0 & \text{elsewhere.} \end{cases}$$

Determine **a** the value of k, **b** $P\left(\frac{\pi}{4} \le X \le \frac{3\pi}{4}\right)$.

18 A particular industrial process involves the extraction of a valuable metal from rock deposits containing the metal. The rock is mined and then processed in an extraction unit in five-tonne 'batches', with each batch containing approximately 100 kg of the metal. The cost of mining and then extracting x kg of the metal from each five tonne batch is $\$C$ where

$$C \approx 25\,000 - 20\,000\log_e\left(1 - \frac{x}{100}\right), \quad x < 100.$$

The company carrying out this mining and extraction process has a contract with a buyer who will buy each kilogram of the extracted metal for $1000.

a Write down an expression for $P(x)$, the profit function.

b (For this part, first determine your answers using calculus and then view the graph of $P(x)$ on a calculator and determine the answers from the graph.)
How many kilograms should be extracted from each five-tonne batch for maximum profit, and what would this maximum profit be?

19 The continuous random variable X has the cumulative distribution function

$$P(X \le x) = \begin{cases} 0 & \text{for} \quad x < 1 \\ \dfrac{x^2 + 3x - 4}{36} & \text{for} \quad 1 \le x \le 5 \\ 1 & \text{for} \quad x > 5. \end{cases}$$

Determine: **a** $P(X \le 2)$ **b** $P(X \ge 2)$
c $P(3 \le X \le 5)$ **d** $P(X > 7)$

20 Let us suppose that the number of kilometres a new tyre of a particular brand lasts before it needs to be replaced is normally distributed with a mean of 60 000 km and a standard deviation of 8000 km.

a Determine the probability that a randomly selected new tyre of this brand lasts for less than 45 000 km before it needs to be replaced.

b Determine the probability that when four new tyres of this brand are randomly selected at least one will need to be replaced before it has travelled 45 000 km.

iStock.com/algre

ISBN 9780170395205