Shutterstock.com/bikeriderlondon

8

CONTINUOUS RANDOM SAMPLES AND THE NORMAL DISTRIBUTION

CONTINUOUS RANDOM SAMPLES AND THE NORMAL DISTRIBUTION

TERMINOLOGY

closed interval
continuous
cumulative distribution function (cdf)
discrete
expected value
linear transformation
mean
median
normal distribution
open interval
probability density function (pdf)
quantile
quartile
random variable
range
rectangular distribution
semi-closed interval
standard deviation
standard normal distribution
standard normal score
triangular distribution
uniform distribution
variance
Z-score

Prior learning

GENERAL CONTINUOUS RANDOM VARIABLES

- use relative frequencies and histograms obtained from data to estimate probabilities associated with a continuous random variable (ACMMM164)
- understand the concepts of a probability density function, cumulative distribution function, and probabilities associated with a continuous random variable given by integrals; examine simple examples of continuous random variables and use them in appropriate contexts (ACMMM165)
- recognise the expected value, variance and standard deviation of a continuous random variable and evaluate them in simple cases (ACMMM166)
- understand the effects of linear changes of scale and origin on the mean and the standard deviation (ACMMM167)

NORMAL DISTRIBUTIONS

- identify contexts such as naturally occurring variation that are suitable for modelling by normal random variables (ACMMM168)
- recognise features of the probability density function of the normal distribution with mean μ and standard deviation σ and the use of the standard normal distribution (ACMMM169)
- calculate probabilities and quantiles associated with a given normal distribution using technology, and use these to solve practical problems (ACMMM170) AC

8.01 CONTINUOUS RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS

You have already studied discrete random variables and distributions. You have also examined the most important of the discrete distributions, the binomial distribution.

In this chapter you will study continuous random variables and the **normal distribution**, which is the most important continuous distribution.

You should already know the difference between a **discrete variable** and a **continuous variable** from your work in earlier years. You have also used **random variables** in Chapters 2 and 5.

IMPORTANT

A **random variable** is a variable with a numerical value that depends on the outcome of a chance experiment. A random variable is **discrete** if the values of the variable (**range**) are either a finite set, the counting numbers or an equivalent set. A random variable is **continuous** if it can take any real value, or any real value from an interval.

A random variable is usually denoted by a capital letter. The corresponding lower-case letter denotes specific values of the variable.

The heights of people selected at random, the time you wait for a bus, the mass of a pebble selected at random from some gravel and length of a randomly selected city street are all examples of continuous random variables whose values are numbers from an interval.

You can use data and relative frequencies to examine the distributions of continuous random variables and estimate associated probabilities. When estimating probabilities across groups, you need to take account of the proportion of each group that is included.

9780170250290

Example 1

Julian timed the duration of his bus trips to and from school each day for a fortnight. the nearest 30 seconds) it took to get from his stop to school and vice versa, including the time at stops along the way. The times in minutes he obtained were as follows:

$17\frac{1}{2}$, $18\frac{1}{2}$, $19\frac{1}{2}$, $17\frac{1}{2}$, 18, 15, $17\frac{1}{2}$, 15, 15, $15\frac{1}{2}$, 19, $16\frac{1}{2}$, 21, $13\frac{1}{2}$, 17, 17, 15, 19, 15, $14\frac{1}{2}$, 16, $14\frac{1}{2}$, $16\frac{1}{2}$, $17\frac{1}{2}$, 16, 14, 16, $19\frac{1}{2}$

a Draw a frequency table and histogram of the times in 1 minute intervals.

b Use the histogram to estimate the probability of Julian's bus trip taking 14–16 minutes, to the nearest minute.

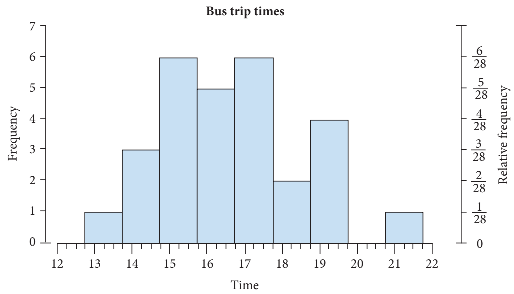
Solution

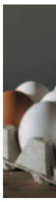
a The times are from 13.5 minutes to 21 minutes. Draw the table for 13–13.5, 14–14.5, … minutes. It is also a good idea to add a relative frequency column to the table.

Time (minutes)	Frequency	Relative frequency
13–13.5	1	$\frac{1}{28}$
14–14.5	3	$\frac{3}{28}$
15–15.5	6	$\frac{3}{14}$
16–16.5	5	$\frac{5}{28}$
17–17.5	6	$\frac{3}{14}$
18–18.5	2	$\frac{1}{14}$
19–19.5	4	$\frac{1}{7}$
20–20.5	0	0
21–21.5	1	$\frac{1}{28}$
Total	28	1

Draw the histogram, with true class intervals from 11.75–12.75, 12.75–13.75, …

You can put a relative frequency on the RHS to make part **b** easier.





b	Write down what parts of the class intervals need to be used for 13.5–16.5 minutes.	13.5–13.75 is 1 of 12.75–13.75, all of 13.75–14.75 and 14.75–15.75 are included and 15.75–16.5 is $\frac{3}{4}$ of 15.75–16.75.
	Find the probability for 13.5–16.5 minutes.	$P(14\text{–}16)=\frac{1}{4}\times\frac{1}{28}+1\times\frac{3}{28}+1\times\frac{6}{28}+\frac{3}{4}\times\frac{5}{28}$ $=\frac{13}{28}\approx 0.46$
	Write the answer.	The estimated probability of the trip taking between 14–16 minutes, to the nearest minute, is about 46%.

In fact, since the universe itself is finite, there cannot be a real frequency distribution that is on the whole set of real numbers, but for practical purposes there are many that are effectively on the whole set of real numbers. Example **1** showed a practical example of a continuous random variable on a small interval. Although the data was collected discretely, it is still actually continuous. This is an important distinction, because it affects the mathematical treatment of the information.

WS
Discrete and continuous random variables

Some continuous random variables are on intervals that include negative numbers, others start at 0, but many are similar to Example **1**, having intervals between positive bounds. The class interval 5–10 usually means $5 \le x < 10$, but 5–10 inclusive means $5 \le x \le 10$. Strictly between 5 and 10 means $5 < x < 10$.

Example 2

The thickness of a layer of oil on top of some water depends on the amount of oil that is used, the viscosity of the oil, the temperature and the time it is allowed to spread. The thickness can be measured by observing the way angled light is reflected from the oil/water. A researcher made the following observations of thicknesses for different oils at temperatures like those at coastal locations. The thicknesses are in millimetres.

1.62, 0.71, 0.97, 0.04, 1.87, 1.22, 1.39, 0.34, 2.14, 1.27, 2.09, 0.49, 0.31, 2.47, 1.71, 0.91, 2.57, 0.51, 0.67, 0.61, 0.68, 0.74, 0.63, 0.04, 0.46, 2.48, 2.31, 1.69, 0.43, 2.29, 1.88, 2.53, 1.35, 0.43, 1.86

a Draw a histogram of the thicknesses with class widths of 0.25.
b Find the probability that the thickness of a randomly selected sample is strictly between 0.9 and 1.4 mm.

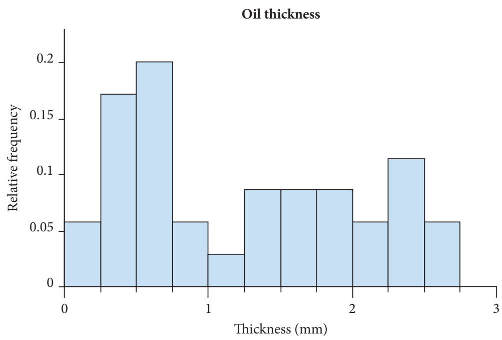
Solution

a Draw a table first. The thickness cannot be less than 0.

Thickness	Frequency	R. F.
0–0.25	2	0.057
0.25–0.5	6	0.171
0.5–0.75	7	0.2
0.75–1	2	0.057
1–1.25	1	0.029
1.25–1.5	3	0.086
1.5–1.75	3	0.086
1.75–2	3	0.086
2–2.25	2	0.057
2.25–2.5	4	0.114
2.5–2.75	2	0.057
Totals	35	1.000

9780170250290

Now draw the histogram. In this case, the true class intervals are almost the same as the stated class intervals. Use the relative frequency as the vertical scale.



b Write down what parts of the class intervals need to be used.

0.9–1 is $\frac{2}{5}$ of 0.75–1, all of 1–1.25 is included and 1.25–1.4 is $\frac{3}{5}$ of 1.25–1.5.

Find the probability for 0.9–1.4 mm.

$$P(0.9\text{–}1.4) = \frac{2}{5} \times 0.057 + 1 \times 0.029 + \frac{3}{5} \times 0.086$$
$$= \frac{18}{56} \approx 0.103$$

Write the answer.

The estimated probability of the oil thickness being strictly between 0.9 and 1.4 mm is about 0.103.

INVESTIGATION Belly button heights

The belly button heights of some Australian Year 12 students who did the Census in schools survey are as follows, showing gender and height in cm.

M, 120; M, 100; M, 106; F, 109; F, 100; M, 110; F, 92; F, 110; M, 95; F, 108; M, 112; F, 119; M, 103; M, 105; F, 94; F, 91; M, 120; F, 100; F, 102; F, 105; F, 100; M, 113; M, 95; F, 97; F, 105; F, 104; F, 100; M, 115; M, 82; M, 108; F, 94; M, 100; M, 100; F, 106; M, 100; M, 110; F, 100; F, 112; F, 80; F, 100; F, 105; F, 100; M, 110; F, 97; F, 108; F, 100; F, 101; F, 100; M, 109; M, 100; F, 94; F, 97; F, 93; M, 106; M, 105

- Construct a histogram of the data.
- Find the probability of a randomly selected Australian student having a BB height within 3 cm of your own.
- Find the mean and standard deviation of the data.
- Collect the data for your class.
- Compare the data for your class with the data for Australian students in general.
- What do you notice about the patterns of the data?

EXERCISE 8.01 Continuous random variables and probability distributions

Concepts and techniques

1 Examples 1, 2 The data below shows the palm widths, to the nearest 0.1 cm, of a group of Year 12 students.

7.2, 8.2, 10.2, 6.4, 6.6, 5.7, 6.4, 5.7, 9.1, 8.8, 11.1, 6.8, 8.8, 9.1, 7.5, 6.5, 9.4, 8.8, 9.3, 7.1, 9.9, 8.9, 5.8, 10, 9.2, 8.3, 11.2, 9, 7.8

a Draw up a frequency table with classes 5.5–5.9, 6.0–6.4, 6.5–6.9, …
b Construct a relative frequency histogram.
c Use the histogram to estimate the probability of a palm width of strictly between 6 and 7 cm?
d Use the histogram to estimate the probability of a palm width of 8 cm, measured to the nearest cm?

2 The data below shows the heights of a group of 17-year-olds to the nearest cm.

162, 161, 158, 169, 161, 180, 166, 159, 171, 163, 166, 163, 181, 170, 178, 175, 164, 169, 174, 160

a Make a relative frequency table with class widths of 5 cm.
b Draw a relative frequency histogram
c Use the histogram to estimate the probability of a randomly selected 17-year-old having a height of 161–164 cm.
d Use the histogram to estimate the probability of having a height over 168 cm?
e Use the histogram to estimate the probability of having a height strictly between 165 and 175 cm?

3 Sarah has kept track of the amount of time she waits to get served when she goes to the post office with the afternoon mail. The times to the nearest minute are shown below.

6, 7, 4, 9, 0, 6, 4, 9, 1, 6, 1, 7, 2, 7, 2, 10, 2, 9, 1, 9, 0, 6, 6, 9, 6, 8, 4, 12

a Make a frequency table with class widths of 2 minutes.
b Draw a relative frequency histogram.
c Estimate the probability of waiting less than 5 minutes.
d Estimate the probability of waiting more than 10 minutes?
e Estimate the probability of waiting strictly between 5 and 10 minutes?

Reasoning and communication

4 The following are the IQs of a group of Year 12 students. IQ is a continuous variable.

90, 111, 89, 123, 137, 97, 120, 101, 101, 101, 110, 108, 101, 85, 103, 121, 92, 130, 109, 112, 121, 139, 100, 83, 141, 82, 94, 94, 119, 104, 95, 87, 108, 104, 101, 115, 106, 93, 90, 128, 107, 110, 98, 138

a Draw up a frequency table with class widths of 10 and construct a histogram.
b What is the probability of an IQ of 100–110?
c What is the probability of an IQ strictly between 100 and 110?
d What is the mean IQ?
e IQ tests are standardised to have a mean of 100. Why isn't the mean of this group 100?

5 Said waits for a tram to go to work in the city each morning. He walks across the King's domain to St Kilda Rd but often stops for a chat on the way. The times in minutes (to the nearest 30 seconds) that he takes to get to the tram stop are given below.

Alamy/William Caram

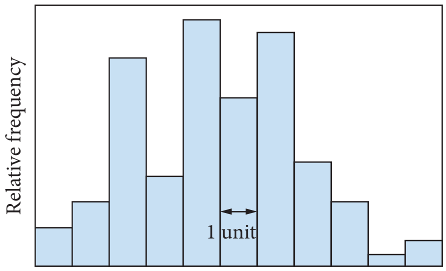
2, 7.5, 5.5, 3.5, 1, 4.5, 2.5, 12.5, 11.5, 1.5, 3.5, 0, 3.5, 7, 3.5, 10.5, 8.5, 7.5, 7.5, 4.5, 9.5, 10.5, 12, 15.5, 0.5, 10.5, 0, 1, 2.5, 4, 8, 4.5, 2.5, 7.5, 1.5

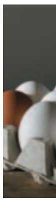
a Draw up a frequency table with classes of width 2 minutes and construct a histogram.
b Estimate the probability of waiting for strictly less than 5 minutes?
c Can you make any conclusion about the trams going along St Kilda Rd in the morning?

8.02 PROBABILITY DENSITY AND CUMULATIVE DISTRIBUTION FUNCTIONS

The probability of you getting *an exact value* of a continuous random variable is theoretically zero because it is one value from an infinite number of values. In practice, continuous quantities are measurements, and they can never be exact because anything you use for measurement has limited accuracy. As in the previous section, you will want to know the probability of a continuous random variable being between certain limits. Even if you ask 'what is the probability of someone weighing 70 kg?', you really mean a value from 69.5 kg to 70.5 kg, unless you state greater or less accuracy.

Consider the following histogram of relative frequencies worked out from data for class widths of 1 unit.





Each column must be of area equal to the relative frequency. Since the relative frequencies add up to 1, this means that the area of the histogram must also be 1. You work out the probability of getting between particular scores by finding the area of that part of the histogram.

For very large collections of data, you will have more columns and the graph will start to look like a line graph, but the area under the line will still be 1. The probability of getting between particular scores will still be the area under the curve.

IMPORTANT

The **probability density function** for a continuous random variable X is defined as a function $p(x)$ such that $P(c \le X \le d) = \int_c^d p(x)dx$ for all intervals (c, d) on which X is defined.

The probability density function is often referred to by its initials **pdf**.

You should be able to see that a probability density function cannot have negative values because this would give a negative probability. Because the area under the function is a probability, the total area under the function must be 1.

IMPORTANT

A probability density function $p(x)$ for a continuous random variable X defined on the interval $[a, b]$ must satisfy the conditions $p(x) \ge 0$ for $x \in [a, b]$ and $\int_a^b p(x)dx = 1$.

The **closed interval** $a \le x \le b$ includes the end values a and b and is written as $[a, b]$.

The **open interval** $a < x < b$ excludes the end values a and b and is written as (a, b).

The semi-closed interval $a \le x < b$ includes a and excludes b and is written as $[a, b)$.

The real numbers can be written as $(-\infty, \infty)$. It is conventional to exclude 'infinity' because the real numbers are unbounded in either direction, so you cannot reach the end.

Example 3

For each of the following functions, determine whether or not it could be a probability density function on the given interval.

a $f(x) = \frac{3}{32}(4 - x^2)$ for the interval $[-2, 2]$

b $f(x) = \begin{cases} 3x^2 & \text{for } 0 \le x \le 1 \\ 0 & \text{for } -\infty < x < 0 \text{ or } x > 1 \end{cases}$

c $f(x) = x$ for the interval $[0, 3]$

d $f(x) = \dfrac{1}{2(|x|+1)^2}$ for the real numbers

Solution

a Check that $f(x) \ge 0$.

$|x| \le 2$ on $[-2, 2]$, so $4 - x^2 \ge 0$ and $f(x) \ge 0$

Find $\int_{-2}^{2} f(x)dx$.

$$\int_{-2}^{2} \frac{3}{32}(4 - x^2)dx = \frac{3}{32}\left[4x - \frac{x^3}{3}\right]_{-2}^{2}$$
$$= \frac{3}{32}\left[\left(8 - \frac{8}{3}\right) - \left(-8 - \frac{-8}{3}\right)\right]$$
$$= \frac{3}{32}\left[\frac{48 - 16}{3}\right]$$
$$= 1$$

Write the conclusion.

Since $f(x) \geq 0$ and $\int_{-2}^{2} f(x)dx = 1$, $f(x)$ is a pdf on $[-2, 2]$.

b Check that $f(x) \geq 0$.

$3x^2 \geq 0$ so $f(x) \geq 0$

Find $\int_0^1 f(x)dx$.

$$\int_0^1 3x^2 dx = \left[x^3\right]_0^1 = 1$$

Write the conclusion.

Since $f(x) \geq 0$ and $\int_0^1 f(x)dx = 1$,

$$f(x) = \begin{cases} 3x^2 & \text{for } 0 \leq x \leq 1 \\ 0 & \text{for } -\infty < x < 0 \text{ or } x > 1 \end{cases}$$

is a pdf

c Check that $f(x) \geq 0$.

$x \geq 0$ for $x \in [0, 3]$ so $f(x) \geq 0$

Find $\int_0^3 f(x)dx$.

$$\int_0^3 x\,dx = \left[\frac{x^2}{2}\right]_0^3$$
$$= \frac{9}{2} - 0$$
$$= 4\frac{1}{2}$$

Write the conclusion.

Since $\int_0^3 f(x)dx \neq 1$, $f(x)$ is not a pdf on $[0, 3]$.

d Check that $f(x) \geq 0$.

$\dfrac{1}{2(|x|+1)^2} \geq 0$ for all $x \in \boldsymbol{R}$ so $f(x) \geq 0$.

Find $\int_{-\infty}^{\infty} f(x)dx$.

$$\int_{-\infty}^{\infty} \frac{dx}{2(|x|+1)^2} = \int_{-\infty}^{0} \frac{dx}{2(|x|+1)^2} + \int_0^{\infty} \frac{dx}{2(|x|+1)^2}$$
$$= \left[-\frac{-1}{2(1-x)}\right]_{-\infty}^{0} + \left[\frac{-1}{2(x+1)}\right]_0^{\infty}$$
$$= \left(\frac{1}{2} - 0\right) + \left(0 - \frac{-1}{2}\right)$$
$$= 1$$

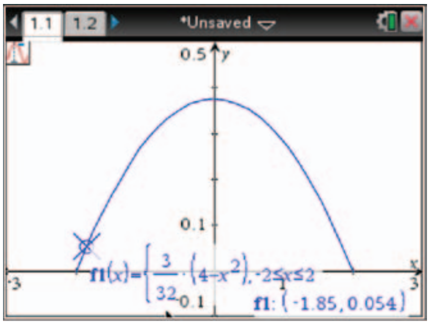
Write the conclusion.

Since $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)dx = 1$,

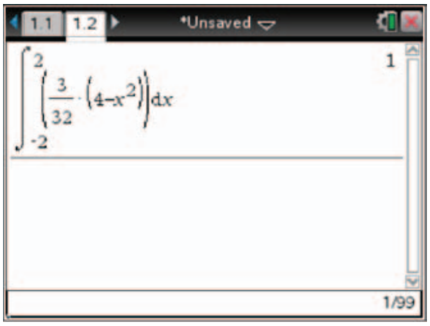
$f(x) = \dfrac{1}{2(|x|+1)^2}$ is a pdf on $\boldsymbol{R}$.

TI-Nspire CAS

a Use the graph page to draw the function for the interval $-2 \leq x \leq 2$ to check the sign. Use ctrl = to get the signs and use Trace and Trace step to check the whole graph.

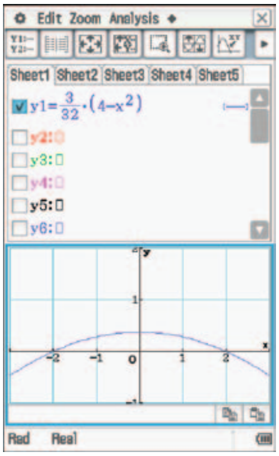


Then use the calculator page to find the integral.

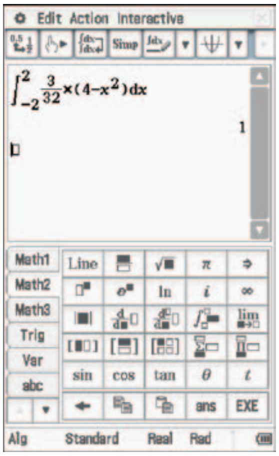


ClassPad

a Use the Graph & Table menu to draw the function for the interval $-3 \le x \le 3$ to check the sign. Use View Window to set $-2 \le x \le 2$ and $-1 \le y \le 2$.



Use the Main menu and (Keyboard) Math2 to calculate the integral. Integrals are found in the Math2 menu. Make sure the calculator is set to Standard.



The area under a curve for the interval (a, b) is the same as the area under the curve for the closed interval $[a, b]$. This means that for a continuous random variable the probabilities $P(a \le X \le b)$, $P(a < X \le b)$, $P(a \le X < b)$ and $P(a < X < b)$ are all the same.

You can make a probability density function from any continuous function on an interval where it is positive by dividing the function by its integral over the interval. It will only be useful when it corresponds to a probability situation.

Example 4

Make a probability density function using $x^2 - x + 2$ on the interval [−3, 4].

Solution

Check that the function is positive on the interval.

$$x^2 - x + 2 = \left(x - \frac{1}{2}\right)^2 + 1\frac{3}{4} > 0 \text{ for all } x \in \boldsymbol{R}$$

Find $\int_{-3}^{4} (x^2 - x + 2)dx$.

$$\int_{-3}^{4} (x^2 - x + 2)dx = \left[\frac{x^3}{3} - \frac{x^2}{2} + 2x\right]_{-3}^{4}$$

$$= \left(\frac{64}{3} - \frac{16}{2} + 8\right) - \left(\frac{-27}{3} - \frac{9}{2} - 6\right)$$

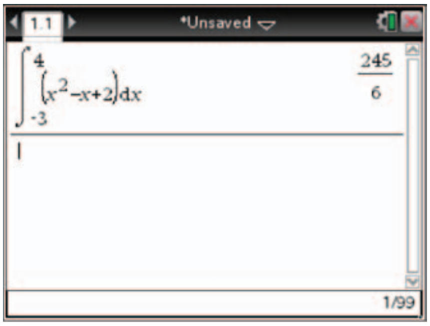
$$= \frac{245}{6}$$

Make a function such that $\int_{-3}^{4} (x^2 - x + 2)dx = 1$.

$f(x) = \dfrac{6(x^2 - x + 2)}{245}$ is a pdf on [−3, 4].

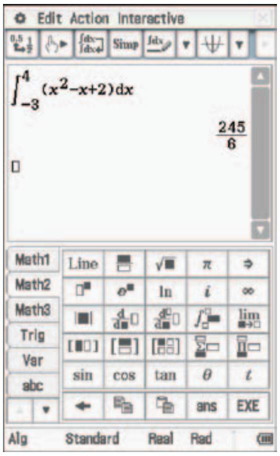
TI-Nspire CAS

You can find the integral using your CAS calculator.



ClassPad

Make sure that the calculator is set to Standard. Use the Main menu and the Math2 menu.



You can use probability density functions to work out probabilities associated with continuous random variables by integrating the function, but it is more useful to use a function that gives the probability directly.

IMPORTANT

The **cumulative distribution function** for the probability density function $f(x)$ of a continuous random variable X defined on the interval (a, b) is given by $F(x) = \int_a^x f(x)dx$.

The cumulative distribution function is often referred to by its initials **cdf**.

It follows immediately that $P(c \le x \le d) = F(d) - F(c)$.

Example 5

a Find the cumulative distribution function for the probability density function $f(X = x) = x^2$ defined on the interval $\left[0, \sqrt[3]{3}\right]$.

b Use the cdf to find the probability that $0.7 < X < 1.3$.

Solution

a Check that f is a pdf.

$$\int_0^{\sqrt[3]{3}} x^2 dx = \left[\frac{x^3}{3}\right]_0^{\sqrt[3]{3}} = 1$$

State the result.

Since $x^2 \ge 0$ as well, it is a pdf.

Find the integral.

$$F(X) = \int_0^x x^2 dx = \left[\frac{x^3}{3}\right]_0^x = \frac{x^3}{3}$$

b Write $P(a < X < b)$ as the difference of the values of $F(x)$.

$P(0.7 < X < 1.3) = F(1.3) - F(0.7)$

Calculate the value.

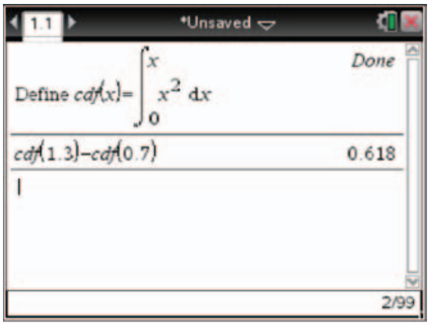
$$= \frac{(1.3)^3}{3} - \frac{(0.7)^3}{3} = 0.618$$

Write the answer.

The probability is 0.618.

TI-Nspire CAS

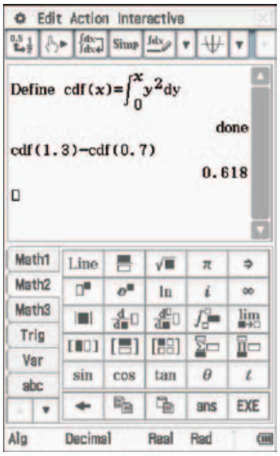
You can define the cdf using the integral on your CAS calculator.



9780170250290

ClassPad

Note that $\int_0^x x^2 dx$ does not really make sense, as it literally means the integral from $x = 0$ to $x = x$. The TI-Nspire seems to be able to cope with this, but you must use two variables for the ClassPad.



EXERCISE 8.02 Probability density and cumulative distribution functions

Concepts and techniques

Probability density functions

1 Example 3 For each of the following functions, determine whether or not it could be a probability density function on the given interval.

a $f(x) = 0.5x$ for the interval [0, 2]

b $f(x) = \dfrac{1}{(x+1)^2}$ for the interval $[0, \infty)$

c $f(x) = \frac{1}{3}(3 - x)(x + 1)$ for the interval [0, 3]

d CAS $f(x) = 4x^3 - 4x$ for the interval $\left[-1, \sqrt{2}\right]$

e CAS $f(x) = \dfrac{e}{7x}$ for the interval [1, 7]

2 Example 4 Make probability density functions from each of the following functions on the given intervals.

a $f(x) = \dfrac{1}{(x-1)^2}$ for the interval $[5, \infty)$

b $f(x) = x^3$ for the interval [0, 4]

c CAS $f(x) = (3 - x)(x + 3)$ for the interval [−1, 2]

d CAS $f(x) = x$ for the interval [0, 20]

e CAS $f(x) = e^{2x} - 1$ for the interval [0, ln(2)]

3 Example 5 a Find the cumulative distribution function for the probability density function $f(x) = x^{-2}$ defined on the interval $[1, \infty)$.

b Use the cdf to find the probability that $1 < X < 2$.

c Use the cdf to find the probability that $2 < X < 3$.

d Use the cdf to find the probability that $2 < X < 4$.

e Use the answers to **c** and **d** to find the probability that $3 < X < 4$.

4 CAS a Find the cumulative distribution function for the probability density function $f(x) = \frac{1}{70}(x+2)$ defined on the interval [0, 10].

b Use the cdf to find the probability that $0.5 < X < 1$.

c Use the cdf to find the probability that $2 < X < 4$.

d Use the cdf to find the probability that $5 < X < 10$.

e Use the cdf to find the probability that $4 < X < 6$.

5 CAS a Find the cumulative distribution function for the probability density function $f(x) = e^{-x}$ defined on the interval $[-\ln(1.2), \ln(5)]$.

b Use the cdf to find the probability that $2 < X < 3$.

c Use the cdf to find the probability that $-0.1 < X < 1$.

d Use the cdf to find the probability that $0 < X < 4$.

e Use the cdf to find the probability that $-0.15 < X < 0.15$.

Reasoning and communication

6 The cumulative distribution function for the continuous random variable Q defined for the domain [0, 5] is given by $F(x) = 0.4x - 0.04x^2$. Find the probability density function $f(x)$, describe the shape of $f(x)$ and prove that it is actually a probability density function.

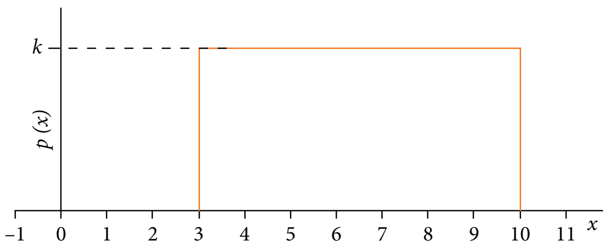
7 A cumulative distribution function is given by $F(t) = 1 - e^{-4t}$ on the domain $[0, \infty)$. Find the probability density function $f(t)$, describe its shape and prove that it is actually a probability density function.

8 CAS A cumulative distribution function is given by $F(m) = 0.1m$ for the domain [0, 10]. Find the probability density function, describe its shape and prove that it is actually a probability density function.

8.03 SIMPLE CONTINUOUS RANDOM VARIABLES

The simplest functions are constants and straight lines. These are also the simplest probability density functions that have practical applications.

Consider a constant probability density function $p(x) = k$ on the interval [3, 10]. Can you work out the value of the constant k?



Since $p(x)$ is a probability density function, the total area under the line from 3 to 10 is 1.

It is a rectangle of length $10 - 3 = 7$. You should be able to see that the area is $7k$, so $7k = 1$ and $k = \frac{1}{7}$. The same method applies for all constant probability density functions.

9780170250290

IMPORTANT

A **uniform continuous probability variable** is one whose probability density function has a constant value on the domain of X.

If X is defined for the domain $[a, b]$, then $p(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{otherwise (i.e., } x \notin [a,b]) \end{cases}$

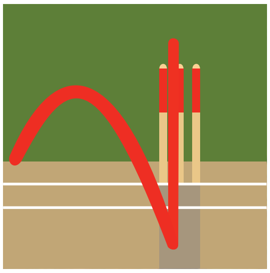
Example 6

A fast bowler aiming for the left stump has equal chances of the ball landing and going anywhere across a 2-foot wide cross-section centred on the outside of the stump. The stumps themselves occupy 9 inches of the cross-section and 1 foot = 12 inches.

a Find the probability density function for X, the distance from the left-hand end of the cross-section that the ball crosses.

b What is the cumulative distribution function for X?

c What is the probability of the ball being in-line with the stumps?



Solution

a The cross-section is 24 inches wide. — X has domain [0, 24].

Write the probability density function. — $p(x) = \dfrac{1}{24}$

b Integrate to find the cumulative distribution function. — $P(x) = \displaystyle\int_0^x \frac{1}{24}\,dx = \frac{x}{24}$

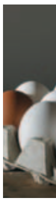
c Where are the stumps? — The stumps occupy the interval [12, 21].

Find the probability that they are in-line. — $P(\text{Stumps in-line}) = P(21) - P(12)$

Calculate the value.

$$= \frac{21}{24} - \frac{12}{24}$$
$$= \frac{9}{24}$$
$$= \frac{3}{8}$$

Write the answer. — The probability that the ball is in-line with the stumps is $\frac{3}{8}$.



A uniform probability distribution is sometimes called a **rectangular distribution**. Could there be a useful distribution with a triangular shape?

IMPORTANT

A **triangular continuous random variable** is one whose probability density function $p(x)$ has a graph in the shape of a triangle.

Example 7

A measurement is reported as being 15 cm ± 0.2 cm.

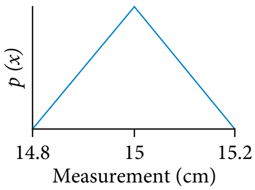
a Draw a triangular graph to show the probability density function.

b Write an expression for the probability density function.

c Find the probability that the measurement is actually 15 cm ± 0.07 cm.

Solution

a The most likely measurement is 15 cm, and the stated error means that the probability of errors diminishes to 0 at 14.8 and 15.2 cm. Draw the graph as a triangle.



b The area $\frac{1}{2}bh$ has to be 1, where $h = p(15)$.

$\frac{1}{2} \times 0.4 \times h = 1$

Solve for the maximum value of $p(x)$.

$h = 5$

Find the slope of the first line.

$m = \frac{5}{0.2} = 25$

Use (14.8, 0) in $y - y_1 = m(x - x_1)$.

$y - 0 = 25(x - 14.8)$

Find the slope of the second line.

$m = \frac{-5}{0.2} = -25$

Use (15.2, 0) in $y - y_1 = m(x - x_1)$.

$y - 0 = -25(x - 15.2)$

Write the probability density function.

$$p(x) = \begin{cases} 25(x - 14.8) & \text{for } 14.8 \le x < 15 \\ -25(x - 15.2) & \text{for } 15 \le x \le 15.2 \end{cases}$$

c There are two parts to integrate.

$P(14.93 \le x < 15) = \int_{14.93}^{15} (25x - 370)\,dx$

Simplify the first integral.

$= 0.288\,75.$

Find the second integral.

$P(15 \le x \le 15.07) = \int_{15}^{15.07} (-25x + 380)\,dx$

Simplify.

$= 0.288\,75$

Write the answer.

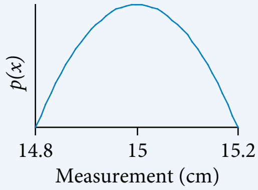
The probability that the measurement is15 cm ± 0.07 cm is $2 \times 0.288\,75 = 0.5775$

9780170250290

In Example **7** the continuous random variable is defined for $[0, \infty)$, but in practical terms it is really $[14.8, 15.2]$. In cases like this where it is obvious that the domain is restricted to a particular interval, the '0 elsewhere' is usually omitted. Notice that the either one or both parts of the domain could include $x = 15$.

INVESTIGATION Continuous probability functions for measurement errors

The main criticism of using a triangular continuous random variable for measurement errors, as shown in Example **8**, is that it has discontinuities at the ends and in the centre. It also gives equal weight to the possible error being anywhere within the limits.



Some researchers prefer to use a quadratic continuous random variable for measurement errors to give greater 'trust' to the central measurement, so their functions look like the diagram given here.

- Assume that the value of the quadratic probability density function at 15 is h
- Write an expression for the function above using the zeros
- Use $(15, h)$ to write $p(x)$ in terms of h
- Find an expression for the area under the curve in terms of h
- Use the fact that the area = 1 to find the function
- Compare the probability that the measurement is actually 15 cm $\pm$ 0.07 cm for a quadratic function with that for a triangular function
- Find other probability density functions used for measurement errors

In some cases, triangular probability density functions have the highest probability at a non-central value.

Example 8

A produce merchant sells chicken feed by the kilogram. Over time, they find that their sales range from 2 kg to 70 kg, with most sales being approximately 20 kg.

a Draw a triangular graph to show the probability density function.

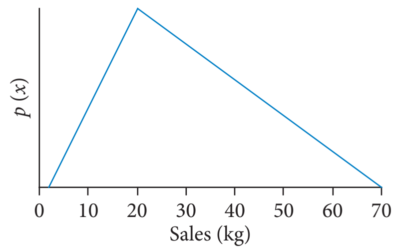
b Write an expression for the probability density function.

c Find the probability that the next sale is between 30 and 40 kg.

Shutterstock.com/Lisa S.

Solution

a The most likely measurement is 20 kg, and the graph has zeros at 2 kg and 70 kg. Draw the graph as a triangle.



b The area $\frac{1}{2}bh$ has to be 1, and $h = p(20)$.

$\frac{1}{2} \times 68 \times h = 1$

Solve for h.

$h = \frac{1}{34}$

Find the slope of the first line.

$m = \frac{\frac{1}{34}}{18} = \frac{1}{612}$.

Use (2, 0) in $y - y_1 = m(x - x_1)$.

$y - 0 = \frac{1}{612}(x - 2)$

Find the slope of the second line.

$m = \frac{-\frac{1}{34}}{50} = -\frac{1}{1700}$

Use (70, 0) in $y - y_1 = m(x - x_1)$.

$y - 0 = -\frac{1}{1700}(x - 70)$

Write the probability density function.

$$p(x) = \begin{cases} \frac{1}{612}(x-2) & \text{for } 2 \le x < 20 \\ -\frac{1}{1700}(x-70) & \text{for } 20 \le x \le 70 \end{cases}$$

c Find the integral.

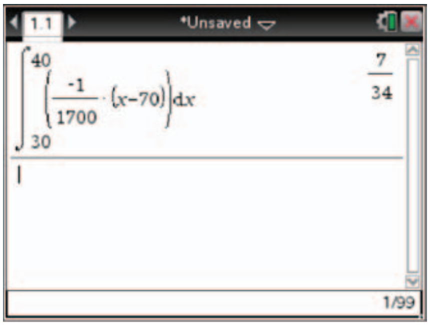
$P(30 \le x < 40) = \int_{30}^{40} -\frac{1}{1700}(x - 70)\,dx$

Calculate the answer.

$= 0.205\ 88$

TI-Nspire CAS

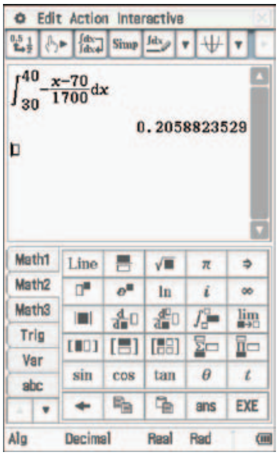
You can use your CAS calculator.



9780170250290

ClassPad

Use CALC from the 2D menu and set the calculator to Decimal.



Write the answer.

The probability that the next sale is between 30 and 40 kg is about 21%.

EXERCISE 8.03 Simple continuous random variables

WS Uniform and triangular probability distribution functions

Concepts and techniques

1 Example 6 Find the uniform probability density functions for continuous random variables defined on the following domains.

a [0, 20] b [0, 18] c [10, 20] d [5, 15] e [6, 36]

2 Example 7 Find the triangular probability density functions with maximum values in the centres for continuous random variables defined on the following domains.

a [0, 12] b [0, 16] c [12, 22] d [4, 24] e [2, 34]

3 Example 8 A triangular probability density function is defined on the domain [4, 10] with a maximum value at $x = 6$.

a What is the maximum value of $p(x)$?

b What is the slope of the line on the left of 6?

c What is the slope of the line on the right of 6?

d Write an expression for the probability density function.

4 Find triangular probability density functions with maximum values as stated for continuous random variables defined on the following domains.

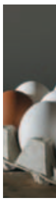
a [5, 15] with a maximum value at 7

b [4, 10] with a maximum value at 8

c [20, 30] with a maximum value at 23

d [0, 20] with a maximum value at 15

e [20, 90] with a maximum value at 60



Reasoning and communication

5 Johannes has found that the time it takes him to drive to work is between 15 and 20 minutes, with equal chances of being any of the times in between.
   a Construct a probability density function $p(t)$ for the time it takes him to get to work.
   b What is the probability of being late if he leaves 16 minutes before he is due at work?
   c What is the probability of being late if he leaves 18 minutes before he is due at work?

6 Buses go past Carol's stop into the city every 10 minutes, so she doesn't bother looking at the timetable before catching the bus.
   a What are the maximum and minimum times she would wait for the next bus?
   b What is the probability of catching a bus between 0 and 5 minutes after getting to the stop?
   c Construct a probability density function $b(t)$ for the probability of waiting t minutes for the next bus when she reaches the stop.
   d What is the probability of Carol getting a bus within 3 minutes of arriving at the stop?

7 The record long jump at a particular high school is 6.84 m, accurate to ±2 cm. Use a triangular probability density function to find the probability that the jump was less than 6.85 m.

8 Bathroom scales are accurate to about ±2% after they have been zeroed. Use a triangular probability density function to find the probability that someone whose bathroom scales show them as weighing 82 kg is between 81 and 83 kg.

9 In the game of darts, the triple 20 can be considered to be almost a rectangle 8 mm high and 17 mm wide. A world-class player can hit within 8 mm laterally from his target point and 12 mm vertically. With his first dart, the player aims for the centre of the triple 20.
   a Draw a probability density function for the vertical distance from the centre.
   b Construct a triangular probability density function for being within d mm vertically from the centre.
   c What is the probability of getting a triple 20 on the first dart?
   d The shaft of the darts is about 8 mm in diameter. If the player does get a triple 20, how will this affect his target area?

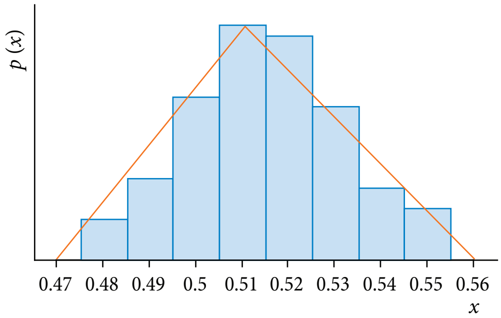
10 Simone has kept note of the time it takes her to get to work. The most common time is 22 minutes, with a minimum of 17 minutes and a maximum of 31 minutes.
   a Construct a triangular probability density function for the time it takes her to get to work.
   b Use your function to find the probability of her taking between strictly between 18 and 20 minutes to get to work.
   c What is the probability of Simone taking 25 minutes to get to work (correct to the nearest minute)?
   d What is the probability of her taking under 25 minutes to get to work?
   e How much time should she allow to get to work?

8.04 EXPECTED VALUE

Front-end loaders are used to load trucks and trailers at a landscape supply business. The bucket on the loader has a nominal size of 0.5 m^3. The actual amount in a bucket was checked 100 times by weight with dry sand, to the nearest 0.002 m^3. The smallest amount was 0.48 m^3, the most common amount

9780170250290

was 0.51 m^3 and the maximum was 0.55 m^3. The amounts were rounded to the nearest 0.01 m^3 for a probability histogram drawn with a triangular probability density function on the same graph.



The triangular graph and the histogram are very alike and would both have areas equal to 1.

If the values were discrete, the expected value would be $\sum x \times p(x)$. This is the same as the mean of the values. If the values were rounded to the nearest 0.004 m^3, the graph would have narrower columns and be more like a curve. Since the values are actually continuous, we should use the limit of $\sum x \times p(x)$ as the widths of the columns is decreased. This corresponds to the integral.

IMPORTANT

The **expected value** of a continuous random variable X defined on the interval $[a, b]$ is calculated from the probability density function $p(x)$ as $\mu = E(X) = \int_a^b x \cdot p(x)dx$.

The expected value is the term used for the theoretical average. The actual average for any set of observations is the **mean** calculated from the histogram. The theoretical average corresponds to an infinite number of observations and is the theoretical **mean**.

Example 9

Calculate the expected value for a uniform probability density distribution defined on the interval [5, 25].

Solution

Write the pdf.	$p(x) = 0.05$
Write the formula for the expected value.	$E(X) = \int_a^b x \cdot p(x)dx$
Substitute for $[a, b]$ and $p(x)$.	$= \int_5^{25} x \times 0.05\,dx$
Calculate the answer.	$= 15$
Write the answer.	The expected value is 15.

As you would expect, the expected value for the uniform continuous random variable in Example **9** is the centre of the distribution.

IMPORTANT

The expected value of a uniform continuous random variable defined on the interval $[a, b]$ is given by $E(X) = \frac{1}{2}(a+b)$.

You might expect the same result for a triangular distribution. However, this is only true for a symmetrical distribution.

Example 10

Find the expected value of a continuous random variable with a triangular probability density function defined on the interval [0.48, 0.55] with a maximum value at 0.51.

Solution

Find the maximum value, at $p(0.51)$.

$\frac{1}{2} \times 0.07 \times h = 1$, so $p(0.51) = h = \frac{200}{7}$

Find the slopes of the lines.

$m_L = \frac{\frac{200}{7}}{0.03} = \frac{20\,000}{21}$, $m_R = \frac{-\frac{200}{7}}{0.04} = -\frac{5000}{7}$

Write $p(x)$.

$$p(x) = \begin{cases} \frac{20\,000}{21}(x - 0.48) & \text{for } 0.48 \le x < 0.51 \\ -\frac{5000}{7}(x - 0.55) & \text{for } 0.51 \le x \le 0.55 \end{cases}$$

Find the expected value.

$$E(X) = \int_a^b x \cdot p(x)dx$$

The integral has two parts.

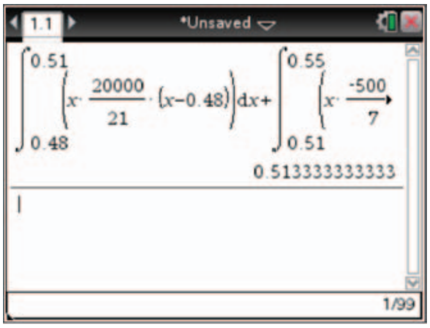
$$= \int_{0.48}^{0.51} x \times \frac{20\,000}{21}(x - 0.48)dx + \int_{0.51}^{0.55} x \times \left[-\frac{5000}{7}(x - 0.55)\right]dx$$

Work out the answer.

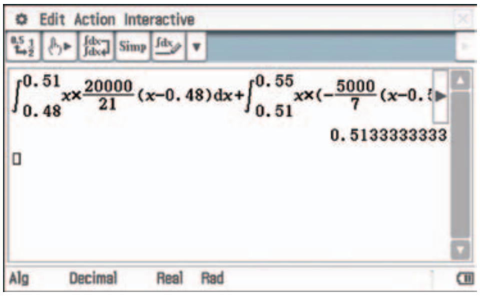
≈ 0.513

You can do this on your CAS calculator.

TI-Nspire CAS



ClassPad



9780170250290

The expected value in Example **10** is a little higher than the point with the maximum value. This corresponds to the mean being higher than the mode in a distribution that is skewed to the right.

EXERCISE 8.04 Expected value

Concepts and techniques

1 Example 9 **a** What is the probability density function for a uniform continuous random variable S on the interval [3, 27]?
b Use integration to find the expected value of S.

2 What is the expected value for a uniform continuous random variable S on the interval [20, 90]?

3 What is the expected value for a uniform continuous random variable S on the interval [8, 36]?

4 Example 10 Find the expected value of a continuous random variable with a triangular probability density function defined on the interval [6, 30] with a maximum value at 24.

5 CAS Find the expected value of a continuous random variable with a triangular probability density function defined on the interval [0, 45] with a maximum value at 10.

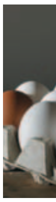
6 CAS Find the expected value of a continuous random variable with a triangular probability density function defined on the interval [20, 60] with a maximum value at 40.

Reasoning and communication

7 Prove that a continuous random variable on $[a, b]$ with a triangular probability density function and maximum in the centre has an expected value of $\frac{a+b}{2}$.

8 CAS The sales of houses in a particular area over a year have a range in prices between \$300 000 and \$750 000, with a mode of about \$450 000.
a Use a triangular distribution function to find the average sale price.
b What is the value of $\int_a^m p(x)dx$ for the median m of a distribution?
c What would you expect the median sale price of houses in this area to be?

8.05 VARIANCE AND STANDARD DEVIATION



In Chapter 5, you saw that the **variance** of a discrete random variable is given by $\sum p(x)(x-\mu)^2$, where μ is the expected value. As with the expected value, the variance for a continuous random variable is worked out using the integral over the interval for which it is defined instead of the sum.

IMPORTANT

The **variance** of a continuous random variable X is calculated using

$$Var(X) = \int_a^b p(x)(x-\mu)^2\,dx$$

where X is defined on $[a, b]$, $p(x)$ is the probability density function and μ is the expected value.

Example 11

Calculate the variance of a uniform continuous random variable defined on the interval [20, 50].

Solution

Find the expected value. $\mu = E(X) = \frac{1}{2}(a+b) = 35$

Find $p(x)$. $p(x) = \frac{1}{50-20} = \frac{1}{30}$

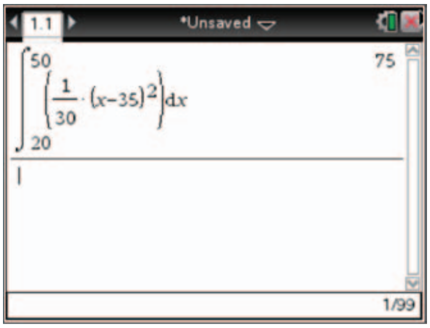
Write the formula. $Var(X) = \int_a^b p(x)(x-\mu)^2\,dx$

Substitute the values. $= \int_{20}^{50} \frac{1}{30}(x-35)^2\,dx$

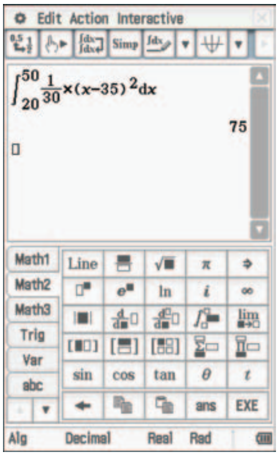
Calculate the value. $= 75$

You can do the integral on your CAS calculator.

TI-Nspire CAS



ClassPad



You can find a formula for the variance of a uniform probability distribution. The **standard deviation** is the square root of the variance.

IMPORTANT

The **standard deviation** of a continuous random variable X is given by $\sigma = \sqrt{Var(X)}$. The standard deviation is also written as $SD(X)$ to emphasise the random variable.

The variance of a uniform continuous random variable X defined on the interval $[a, b]$ is given by $Var(X) = \dfrac{(b-a)^2}{12}$, so the standard deviation is $\dfrac{(b-a)}{2\sqrt{3}}$

Example 12

Find the variance and standard deviation of a continuous random variable X defined on the interval [20, 80] with a symmetrical triangular probability density function.

Solution

The maximum value is at $p(50)$.

$\frac{1}{2} \times 60 \times h = 1$, so $p(50) = h = \frac{1}{30}$

Find the slopes of the lines.

$m_L = \dfrac{\frac{1}{30}}{30} = \dfrac{1}{900}$, $m_R = \dfrac{-\frac{1}{30}}{30} = -\dfrac{1}{900}$

Write $p(x)$.

$$p(x) = \begin{cases} \dfrac{1}{900}(x-20) & \text{for } 20 \le x < 50 \\ -\dfrac{1}{900}(x-80) & \text{for } 50 \le x \le 80 \end{cases}$$

Find the expected value.

By symmetry, $E(X) = 50$

Write the formula.

$$Var(X) = \int_a^b p(x)(x-\mu)^2\,dx$$

Split into the two sections.

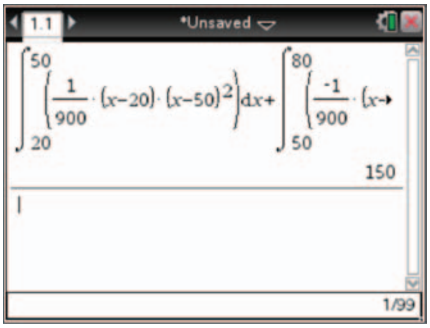
$$= \int_{20}^{50} \frac{1}{900}(x-20)(x-50)^2\,dx + \int_{50}^{80} -\frac{1}{900}(x-80)\,(x-50)^2\,dx$$

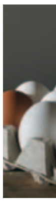
Multiply out to evaluate or use your CAS calculator.

$$= \int_{20}^{50} \frac{1}{900}(x^3 - 120x^2 + 4500x - 50000)dx + \int_{50}^{80} -\frac{1}{900}(x^3 - 180x^2 + 10500x - 200000)dx$$

TI-Nspire CAS

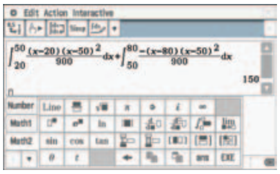
Use the integration function on your CAS calculator for each part.





ClassPad

Use the rotate option Rotate so the screen spreads out lengthways.
Enter the two integrals as a sum.
The calculator should be set to decimal mode.



Write the variance. $Var(X) = 150$

Find the standard deviation. $\sigma = \sqrt{150} \approx 12.25$

You can find a formula for the variance of a continuous random variable with a symmetrical triangular probability density function.

IMPORTANT

The variance of a continuous random variable X defined on the interval $[a, b]$ with a symmetrical triangular probability density function is given by $Var(X) = \frac{(b-a)^2}{24}$, so the standard deviation is $\sigma = \frac{(b-a)}{2\sqrt{6}}$.

EXERCISE 8.05 Variance and standard deviation

Concepts and techniques

1 Example 11 A uniform continuous random variable X is defined on the interval $[4, 16]$.
  a What is the probability density function?
  b What is the expected value?
  c Use calculus to find the variance and standard deviation.

2 Use the formula to find the variance and standard deviation for uniform continuous random variables X defined on the following intervals.
  a $[5, 25]$ b $[0, 50]$ c $[0, 20]$ d $[80, 120]$ e $[0.6, 2.1]$

3 Examples 12 CAS A continuous random variable X is defined on the interval $[12, 20]$ and has a symmetrical triangular probability density function.
  a Find $p(x)$ b Find $E(X)$ c Find $Var(X)$ and σ.

4 CAS A continuous random variable X is defined on the interval $[0, 50]$ and has a symmetrical triangular probability density function.
  a Find $p(x)$ b Find $E(X)$ c Find $Var(X)$ and σ.

5 Use the formula to find the variance and standard deviation for the continuous random variables with symmetrical triangular probability density functions defined on the following intervals.
  a $[5, 15]$ b $[0, 54]$ c $[6, 60]$

6 CAS A continuous random variable X is defined on the interval [0, 10]. It has a triangular probability density function with its maximum value at $x = 0$.

a Find $p(x)$ b Find $E(X)$ c Find $Var(X)$ and σ.

Reasoning and communication

7 Prove that the variance of a continuous random variable X defined on the interval $[a, b]$ with a symmetrical triangular probability density function is given by $Var(X) = \frac{(b-a)^2}{24}$.

8 Find an expression for the variance of a continuous random variable X defined on the interval $[0, a]$ with a symmetrical triangular probability density function that has a maximum at $x = 0$.

8.06 LINEAR CHANGES OF SCALE AND ORIGIN

A linear function is of the form $y = ax + b$.

IMPORTANT

For a random variable X, a **linear change of scale and origin** produces a new random variable $Y = aX + b$, where a and b are constants.

You can use the definitions of expected value, variance and standard deviation to find the effects of a linear change of scale and origin.

Example 13

A uniform continuous random variable X is defined on the interval [24, 144]. It is transformed into the random variable Y according to the equation $Y = 3X - 4$.

a What are the expected value, variance and standard deviation of X?

b What is the interval on which Y is defined?

c What are the expected value, variance and standard deviation of Y?

d What is the relationship between $E(X)$ and $E(Y)$?

e What is the relationship between $Var(X)$ and $Var(Y)$?

Solution

a Use the formula for $E(X)$. $E(X) = \frac{a+b}{2} = \frac{24+144}{2} = 84$

Use the formula for $Var(X)$. $Var(X) = \frac{(b-a)^2}{12} = \frac{(144-24)^2}{12} = 1200$

Now find $SD(X)$. $SD(X) = \sqrt{1200} \approx 34.64$

b Transform 24. For $X = 24$, $Y = 3 \times 24 - 4 = 68$

Transform 144. For $X = 144$, $Y = 3 \times 144 - 4 = 428$

Write the answer. Y is defined on the interval [68, 428].

c	Use the formula for $E(Y)$.	$E(Y)=\dfrac{a+b}{2}=\dfrac{68+428}{2}=248$
	Use the formula for $Var(Y)$.	$Var(Y)=\dfrac{(b-a)^2}{12}=\dfrac{(428-68)^2}{12}=10\,800$
	Now find $SD(Y)$.	$SD(Y)=\sqrt{10\,800}\approx 103.92$
d	Compare $E(X)$ and $E(Y)$.	$3\times E(X)=3\times 84=252$ $E(Y)=3\times E(X)-4$
e	Compare $Var(X)$ and $Var(Y)$.	$9\times Var(X)=9\times 1200=10\,800$ $Var(Y)=3^2\times Var(X)$

Notice that $SD(Y)=3\times SD(X)$ in Example **13**.

Example 14

A continuous random variable X is defined on the interval [30, 90] and has a symmetrical triangular probability density function. It is transformed to the random variable Y according to the equation $Y=0.5X+3$.

a What are the expected value, variance and standard deviation of X?
b What is the interval on which Y is defined?
c What are the expected value, variance and standard deviation of Y?
d What is the relationship between $E(X)$ and $E(Y)$?
e What is the relationship between $Var(X)$ and $Var(Y)$?

Solution

a	Use the formula for $E(X)$.	$E(X)=\dfrac{a+b}{2}=\dfrac{30+90}{2}=60$
	Use the formula for $Var(X)$.	$Var(X)=\dfrac{(b-a)^2}{24}=\dfrac{(90-30)^2}{24}=150$
	Now find $SD(X)$.	$SD(X)=\sqrt{150}\approx 12.25$
b	Transform 30.	For $X=30$, $Y=0.5\times 30+3=18$
	Transform 90.	For $X=90$, $Y=0.5\times 90+3=48$
	Write the answer.	Y is defined on the interval [18, 48].
c	Use the formula for $E(Y)$.	$E(Y)=\dfrac{a+b}{2}=\dfrac{18+48}{2}=33$
	Use the formula for $Var(Y)$.	$Var(Y)=\dfrac{(b-a)^2}{24}=\dfrac{(48-18)^2}{24}=37.5$
	Now find $SD(Y)$.	$SD(Y)=\sqrt{37.5}\approx 6.12$
d	Compare $E(X)$ and $E(Y)$.	$0.5\times E(X)+3=0.5\times 60+3=33=E(Y)$
e	Compare $Var(X)$ and $Var(Y)$.	$0.25\times Var(X)=0.25\times 150=37.5=Var(Y)$ $Var(Y)=(0.5)^2\times Var(X)$

9780170250290

Notice that $SD(Y) = 0.5 \times SD(X)$ in Example **14**.

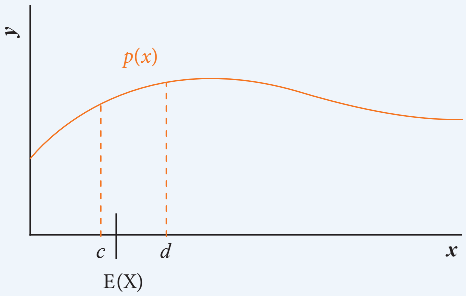
You can apply the methods of Examples **13** and **14** to any uniform or triangular continuous random variables, so the same results will hold for linear transformations of any continuous random variables with uniform or triangular distributions.

INVESTIGATION Linear transformations of continuous random variables

Consider the linear transformation $Y = aX + b$ of some continuous random variable X with probability density function $p(x)$.

Obviously, the graph of probability density function $q(y)$ of the variable Y must be the same basic shape as $p(x)$, but will be altered by the factor a and the shift a along the x-axis.

Consider the part of $p(x)$ from c to d that contains the expected value $E(X)$. It could be shaped as shown below.



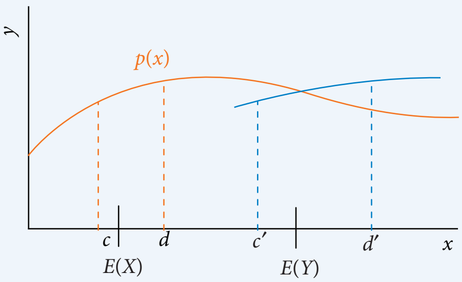
What happens to the interval $[c, d]$ under the linear transformation?

What is c', the value of y corresponding to c under the transformation?

What is d', the value of y corresponding to d under the transformation?

The transformed part of the probability density function could look like the graph below.

The graph is shifted along by the factor b and stretched by the factor a.



Obviously, the probabilities of the two sections must be the same, and the value of $E(Y)$ must be between c' and d'.

$$P(c \le x \le d) = P(c' \le y \le d')$$

For this to be true, what must happen to the vertical scale of the blue shifted version of $p(x)$ so that the areas are the same?

What is the value of x corresponding to a value of y?

Write an expression for the probability density function $q(y)$ of Y in terms of $p(x)$.

You can use the expression for the probability density function of Y from the investigation to find the mean, variance and standard deviation of the transformed random variable.

IMPORTANT

For the linear change of scale and origin of a random variable X with probability density function $p(x)$ given by $Y = aX + b$, where a and b are constants, the probability density function of Y is

$$q(y) = \frac{1}{a}p\left(\frac{y-b}{a}\right),$$
$$E(Y) = aE(X) + b,$$
$$Var(Y) = a^2 Var(X) \text{ and}$$
$$SD(Y) = SD(X)$$

EXERCISE 8.06 Linear changes of scale and origin

Concepts and techniques

1 Example 13 A uniform continuous random variable X is defined on the interval [0, 100]. It is transformed to the random variable Y according to the equation $Y = 2X + 5$.
  a What are the expected value, variance and standard deviation of X?
  b What is the interval on which Y is defined?
  c What are the expected value, variance and standard deviation of Y?
  d What are the relationships between $E(X)$, $Var(X)$, $SD(X)$ and $E(Y)$, $Var(Y)$ and $SD(Y)$?

2 A uniform continuous random variable X is defined on the interval [30, 50]. It is transformed to the random variable Y according to the equation $Y = 5X - 2$.
  a What are the expected value, variance and standard deviation of X?
  b What is the interval on which Y is defined?
  c What are the expected value, variance and standard deviation of Y?
  d What are the relationships between $E(X)$, $Var(X)$, $SD(X)$ and $E(Y)$, $Var(Y)$ and $SD(Y)$?

3 A uniform continuous random variable X is defined on the interval [20, 60]. It is transformed to the random variable Y according to the equation $Y = 0.2X + 4$.
  a What are the expected value, variance and standard deviation of X?
  b What is the interval on which Y is defined?
  c What are the expected value, variance and standard deviation of Y?
  d What are the relationships between $E(X)$, $Var(X)$, $SD(X)$ and $E(Y)$, $Var(Y)$ and $SD(Y)$?

4 Example 14 A continuous random variable X is defined on the interval [7, 8] and has a symmetrical triangular probability density function. It is transformed to the random variable Y according to the equation $Y = 20X - 30$.
  a What are the expected value, variance and standard deviation of X?
  b What is the interval on which Y is defined?
  c What are the expected value, variance and standard deviation of Y?
  d What are the relationships between $E(X)$, $Var(X)$, $SD(X)$ and $E(Y)$, $Var(Y)$ and $SD(Y)$?

9780170250290

5 A continuous random variable X is defined on the interval [0, 50] and has a symmetrical triangular probability density function. It is transformed to the random variable Y according to the equation $Y = 3X - 5$.

a What are the expected value, variance and standard deviation of X?

b What is the interval on which Y is defined?

c What are the expected value, variance and standard deviation of Y?

d What are the relationships between $E(X)$, $Var(X)$, $SD(X)$ and $E(Y)$, $Var(Y)$ and $SD(Y)$?

6 A continuous random variable X is defined on the interval [70, 140] and has a symmetrical triangular probability density function. It is transformed to the random variable Y according to the equation $Y = 0.1X + 2.5$.

a What are the expected value, variance and standard deviation of X?

b What is the interval on which Y is defined?

c What are the expected value, variance and standard deviation of Y?

d What are the relationships between $E(X)$, $Var(X)$, $SD(X)$ and $E(Y)$, $Var(Y)$ and $SD(Y)$?

7 A continuous random variable X is defined on the interval [40, 90] and has an average value of 55 with a standard deviation of 5. It is transformed to the random variable Y according to the equation $Y = 3X + 8$. What are the values of $E(Y)$, $Var(Y)$ and $SD(Y)$?

8 A continuous random variable X is defined on the interval [4, 19]. $E(X) = 14$ and $Var(X) = 8$. X is transformed to the random variable Y according to the equation $Y = 4X - 10$. What are the mean and standard deviation of Y?

Reasoning and communication

9 CAS A continuous random variable X is defined on the interval [0, 25]. It has a triangular probability density function with its maximum value at $x = 0$. It is transformed to the random variable Y by the equation $Y = 8X + 200$.

a Find $p(x)$.

b Find $E(X)$, $Var(X)$ and $SD(X)$.

c What is the interval on which Y is defined?

d Find the probability density function of Y.

e Find $E(Y)$, $Var(Y)$ and $SD(Y)$.

f What are the relationships between $E(X)$, $Var(X)$, $SD(X)$ and $E(Y)$, $Var(Y)$ and $SD(Y)$?

10 CAS A continuous random variable X is defined on the interval [40, 100]. It has a triangular probability density function with its maximum value at $x = 90$. It is transformed to the random variable Y by the equation $Y = 2X - 15$.

a Find $p(x)$.

b Find $E(X)$, $VarX)$ and $SD(X)$.

c What is the interval on which Y is defined?

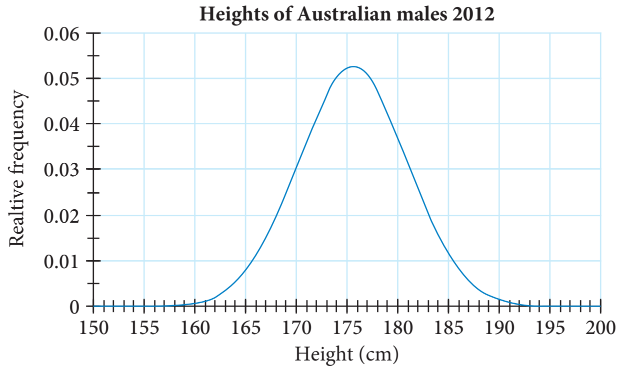
d Find the probability density function of Y.

e Find $E(Y)$, $Var(Y)$ and $SD(Y)$.

f What are the relationships between $E(X)$, $Var(X)$, $SD(X)$ and $E(Y)$, $Var(Y)$ and $SD(Y)$?

8.07 THE NORMAL DISTRIBUTION AND STANDARD NORMAL DISTRIBUTION

The tallest man ever recorded was 272 cm and the shortest man in the world is only 53.5 cm. However, most people are between 120 cm and 210 cm tall. There are healthy people with heights outside this range, as some basketball players are 230 cm, but it is extremely rare. The average height of Australian men in 2012 was found to be 175.6 cm. A relative frequency graph of male height in Australia is shown below.

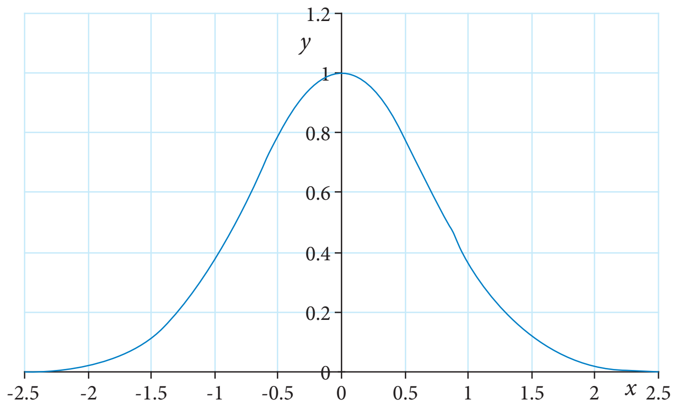


Source: ABS

Although there are adult men in normal health outside the height range shown, the relative frequencies (less than 0.000 03) are too small to be shown on the graph.

A graph like this is often referred to as a 'bell-shaped curve'. It is the most common shape found for continuous random variables that have a central mean and diminishing values to each side, such as human height, IQ, size of raindrops, time taken to complete a task and so on.

The graph of the function $y = e^{-x^2}$ is shown below.



As you can see, the shape of $y = e^{-x^2}$ is identical to the shape of the graph of human height.

9780170250290

It can be shown that $\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$ so the function needs modification to become a probability density function. Writing $p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$ gives a probability density function that has $E(Z) = 0$ and $Var(Z) = 1$, so $SD(Z) = 1$.

The linear transformation $X = \sigma Z + \mu$ changes the distribution to one with mean $E(X) = \mu$, $D(X) = \sigma$, $Var(X) = \sigma^2$ and probability density function $q(x) = \frac{1}{\sigma} p\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, as shown in the last section.

IMPORTANT

The **standard normal distribution** has the probability density function $p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$.

The mean and standard deviation of the standard normal variable Z are $\mu = 0$ and $\sigma = 1$.

A **normal distribution** is a linear transformation of the standard normal distribution with a probability density function of the form $p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, where μ and σ are the mean and standard deviation of the random variable X.

The standard normal distribution and hence any normal distribution assume that the sample space is the set of real numbers, so it is both *continuous* and *infinite*. Notice that the standard normal variable is usually shown as Z.

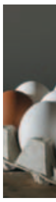
Example 15

What is the probability density function of a normal distribution with $\mu = 100$ and $\sigma = 15$?

Solution

Write the formula.	$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$,
Substitute $\mu = 100$ and $\sigma = 15$.	$= \frac{1}{15\sqrt{2\pi}} e^{-\frac{(x-100)^2}{2\times 15^2}}$
Simplify.	$= 0.0266 e^{-0.002\,222(x-100)^2}$

Since the universe is finite, there are no measurements from an infinite sample space. This means that application of a normal distribution function to practical measurements are necessarily approximations, but they are *very good* ones. Example **15** is the probability density function used for IQ, which is standardised to an average of 100 with a standard deviation of 15, even though scores on the tests used have been rising steadily by the equivalent of about 3 IQ points per ten years.



Example 16

In 2012, the ABS found that the average height of adult Australian females was 161.8 cm with a standard deviation of 6.69 cm.

a Use the normal distribution to model the probability density function of Australian women.

b Use the integral on your CAS calculator to find the probability of an Australian woman being under 152 cm (about 5 feet) tall.

Solution

a Substitute $\mu = 161.8$ and $\sigma = 6.69$ in the formula.

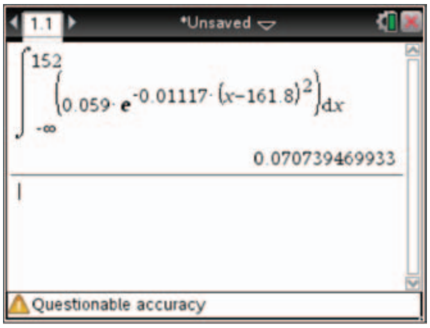
$$p(x) = \frac{1}{6.69\sqrt{2\pi}} e^{-\frac{(x-161.8)^2}{2 \times 6.69^2}}$$

Simplify.

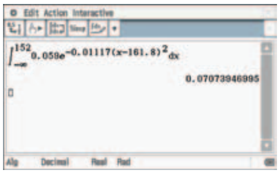
$$= 0.0596e^{-0.01117(x-161.8)^2}$$

b Find $\int_{-\infty}^{152} 0.0596e^{-0.01117(x-161.8)^2}\,dx$.

TI-Nspire CAS



ClassPad



Write the answer.

The probability of an Australian woman being under 152 cm tall is about 7%.

The TI-Nspire gives a warning of 'Questionable accuracy' because the integral cannot be evaluated exactly. The calculator evaluates this integral numerically. In fact, the integral cannot be evaluated algebraically.

As with many complex calculations, tables of values for the standard normal distribution were used until quite recently for calculations normal distributions. The tables showed the areas under the standard normal curve for areas from 0 to Z, where $0 \le Z \le 3$. Since areas outside $-3 \le Z \le 3$ are very small and the curve is symmetrical, this was enough for all but the rarest situations. Modern calculators include normal distribution calculations, so it is unnecessary to use tables.

9780170250290

Example 17

Use your CAS calculator to find the following.

a The area of the standard normal distribution for $-2 < Z \leq 2$.

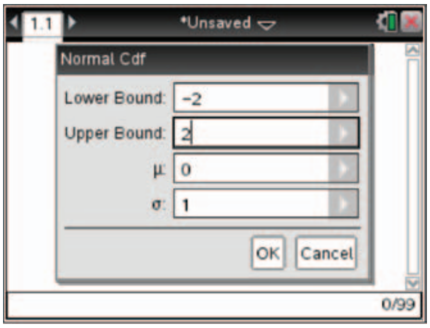
b $P(40 \leq X < 70)$ for the normal variable X with mean 58 and standard deviation 13.

c $P(X > 160)$ for the normal variable X with mean 175 and standard deviation 11.

Solution

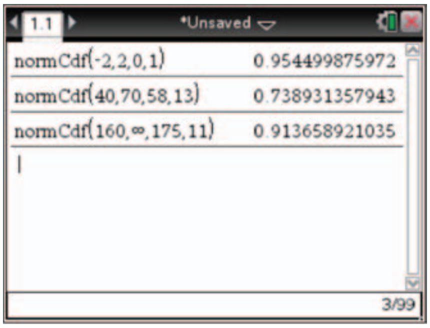
TI-Nspire CAS

a In the **Calculator** ( ).mode, press menu, 6: Statistics ▶, 5: Distributions ▶ and 2: Normal Cdf. Choose −2 for the Lower bound, 2 for the Upper bound, and leave 0 for μ and 1 for σ.



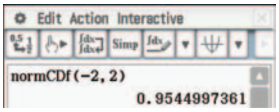
b Choose 40 for the Lower bound, 70 for the Upper bound, and put 58 for μ and 13 for σ.

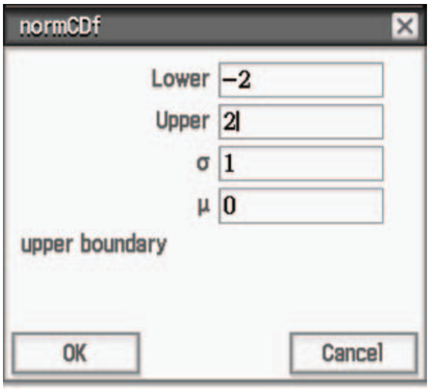
c Use ∞ for the Upper bound.

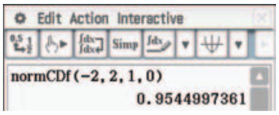


ClassPad

a Tap Action, Distribution/Inv. Dist, Continuous and normCDf. Enter the lower and upper bounds. There is no need to enter a standard deviation or mean as the calculator will assume a standard normal distribution if they are not mentioned. Alternatively tap Interactive, Distribution/ Inv. Dist, Continuous and normCDf. Fill in the table as on the right.

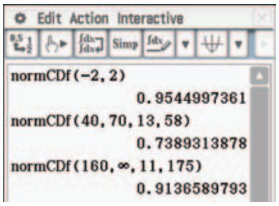






b As for **a**, but add the standard deviation of 13 first, then the mean (58).

c As for **b**, but use ∞ for the upper bound.



Write the answers.

a The area is about 0.9545.
b The probability is about 0.7389.
c The probability is about 0.9137.

What does Example **17** part **a** mean?

For normal distributions, the probability of being within 1 standard deviation of the mean is about 68%, 2 standard deviations is about 95% and 3 standard deviations is about 99.7%. You can check this on your calculator.

WS
The standard normal curve

EXERCISE 8.07 The normal distribution and standard normal distribution

Concepts and techniques

1 Example 15 For each of the following, what is the probability density function for the normal distribution?

a $\mu = 28.5$ and $\sigma = 3.2$
b $\mu = 28.5$ and $\sigma = 5.7$
c $\mu = 48.6$ and $\sigma = 5.7$
d $\mu = 246$ and $\sigma = 78$
e $\mu = 0.07$ and $\sigma = 0.0024$

2 Example 16 CAS For each of the following, state the normal probability density function and find the integral for the stated interval.

a $\mu = 163$ and $\sigma = 8.5$, 150 to 160
b $\mu = 678$ and $\sigma = 147$, 500 to 700
c $\mu = 4240$ and $\sigma = 355$, 4000 to 5000
d $\mu = 6.5$ and $\sigma = 1.8$, 4.9 to 5.8
e $\mu = 74.9$ and $\sigma = 22.6$, 50 to 85

3 Example 17 CAS Find the area of the standard normal distribution table for each of the following intervals.

a $0 \le Z \le 2.3$
b $0 \le Z \le 2.34$
c $0 \le Z \le 1.687$
d $0 \le Z \le 0.058$
e $0 \le Z \le 1.242$

4 CAS Find the following for the standard normal variable Z.

a $P(Z < -1.305)$
b $P(Z > 0.623)$
c $P(Z > 0.596)$
d $P(-1.307 < Z < 2.6)$
e $P(Z < -1.646 \text{ or } Z > 0.831)$

9780170250290

5 CAS Find the following for the standard normal variable Z.

a $P(-1.356 < Z < 1.873)$ b $P(Z < 0.823)$ c $P(Z > -2.306)$

d $P(Z < 1.6)$ e $P(Z < -0.628 \text{ or } Z > 1.636)$

6 CAS For each random normal variable X with the specified mean and standard deviation given below, find the probability shown.

a $\mu = 8463$, $\sigma = 2976$, $P(6766.68 < X < 16766.04)$

b $\mu = 5192$, $\sigma = 2200$, $P(2288 < X < 9196)$

c $\mu = 7.15$, $\sigma = 3.3$, $P(4.807 < X < 14.773)$

d $\mu = 16$, $\sigma = 6$, $P(27.88 < X < 33.04)$

e $\mu = 2.66$, $\sigma = 0.7$, $P(4.095 < X < 5.495)$

7 CAS For each random normal variable X with the specified mean and standard deviation given below, use your CAS calculator to find the probability shown.

a $\mu = 18.33$, $\sigma = 4.68$, $P(20.9508 < X < 28.4388)$

b $\mu = 2175$, $\sigma = 625$, $P(837.5 < X < 3437.5)$

c $\mu = 167.5$, $\sigma = 82.5$, $P(21.475 < X < 491.725)$

d $\mu = 26.98$, $\sigma = 4.94$, $P(30.7344 < X < 45.6532)$

e $\mu = 25.62$, $\sigma = 13.44$, $P(56.6664 < X < 60.1608)$

Reasoning and communication

8 Car tyres have an average life of 50 000 km with a standard deviation of 6150 km. Assuming that the life is normally distributed, find the following.

a The probability of a tyre lasting more than 60 000 km

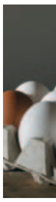
b The probability of a tyre lasting between 45 000 and 55 000 km

c The percentage of tyres that can be expected to last less than 42 000 km

9 A manufacturer of tyres for 4WDs decides to make tyres that last an average of 90 000 km, even though the performance of long-lasting tyres is relatively poor. The standard deviation of the life of these tyres is 8300 km. What percentage of such tyres can be expected to last more than 100 000 km?

10 Over a six-month period, the average exchange rate for the Australian dollar was US\$1.03, with a standard deviation of US\$0.027. The exchange rate is a 'random walk', which means that it is not possible to predict the specific exchange rate at any time and that it is normally distributed. Assume that the pattern is maintained to find the probability that the Australian dollar is worth less than one US dollar.

11 Red wrigglers are an average of 4 cm long, with a standard deviation of 1.2 cm. What is the probability that one of these worms is longer than 5 cm?



12 The average queue time to speak to someone at a bank is 6 minutes, with a standard deviation of 1.5 minutes. What proportion of callers are answered within 3 minutes of calling?

8.08 STANDARDISATION AND QUANTILES

You only need the mean and standard deviation to define a normal distribution. Changes of scale and origin preserve the basic shape and characteristics of a distribution. This means that you can change any normal distribution to any other by using suitable changes to the scale and origin. It is common practice to compare results from different distributions by standardising a score of x to the **standard normal score** z using the linear transformation $Z = \frac{X - \mu}{\sigma}$, where μ and σ are the mean and standard deviation of the normal variable X. The standard normal score is often called the ***Z*-score**.

Example 18

Antoine had a score of 65% for English but only 57% for Physics. The class mean and standard deviation for English were 62% and 8% respectively; for Physics they were 52% and 10%. In which subject did Antoine appear to do better?

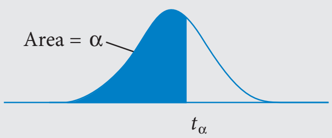
Solution

Calculate the standard normal score for English.	$Z(\text{English}) = \frac{65-62}{8}$ $= 0.375$
Calculate the standard normal score for Physics.	$Z(\text{Physics}) = \frac{57-52}{10}$ $= 0.5$
Write the answer.	Assuming that the classes are of equal ability, Antoine appeared to do better in Physics.

The median divides statistical distributions in half. Quartiles divide distributions into quarters. Percentiles divide distributions into 100 equal groups. **Quantiles** are extensions that divide distributions into proportions of any size.

IMPORTANT

A **quantile** t_α for a continuous random variable X is the value of x such that $P(X < t_\alpha) = \alpha$, where $0 < \alpha < 1$. It is the value that divides the distribution in the ratio $\alpha : 1 - \alpha$.



The **median** is the value m such that $\alpha = 0.5$: $P(X < m) = 0.5$.

You will already be familiar with quartiles. For a continuous random variable, the **first quartile** is the quantile for $\alpha = 0.25$, so a quarter of the distribution lies below it. To find quartiles of normal distributions, you need to use the *inverse* of the cumulative normal distribution function. This function calculates the quantile t_α.

Example 19

V is a random normal variable with mean \$420 000 and standard deviation \$28 000.

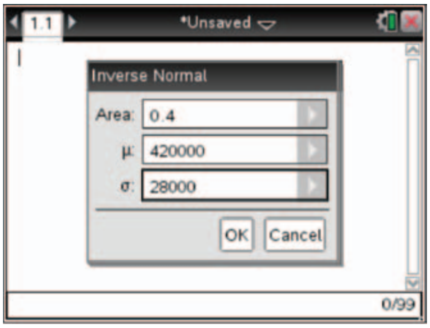
a Find the 40% quantile, the value which 40% of the distribution is below.

b Find the value a such that $P(v > a) = 0.27$.

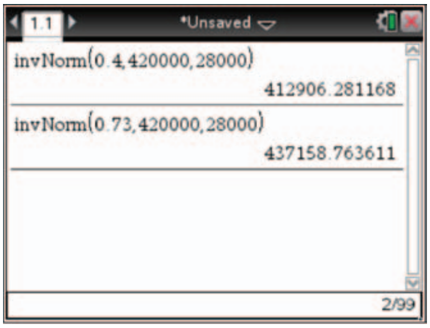
Solution

TI-Nspire CAS

a You need to find $t_{0.4}$.
In the **Calculator** ( ).mode, press menu, 6: Statistics ▶, 5: Distributions ▶ and 3: Inverse Normal. Put 0.4, 420 000 and 28 000 for the Area, μ and σ respectively.

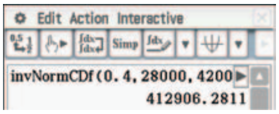


b You want to find the quantile that divides the distribution in the ratio 0.73 : 0.27, so you want $t_{0.73}$.

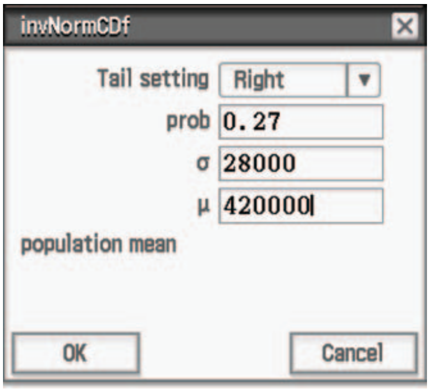


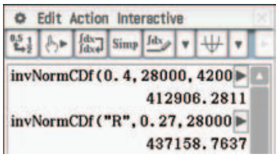
ClassPad

a You need to find $t_{0.4}$.
Tap Action, Distribution/Inv. Dist, Inverse and invNormCDf.
Enter the area, followed by the standard deviation and the mean.
Alternatively, tap Interactive and the same menus and use the table, with the tail setting on Left.



b Tap Interactive, Distribution/Inv. Dist, Inverse and invNormCDf.
Set the Tail to Right, the area is the probability, 0.27, and standard deviation and mean are the same as for part **a**. You can also use the Action menu, but insert the "R" to show that it's a right tail.
Note: Tail setting can also be Center. In this case it assumes the middle area, with both tails cut off symmetrically. In this case use "C", and only the left bound is given.





Write the answers.

a The 40% quantile is about \$413 000.

b The amount which values have a probability of 0.27 of being above is about \$437 000.

EXERCISE 8.08 Standardisation and quantiles

Concepts and techniques

1 Example 19 CAS A random normal variable X has a mean of 10.6 and a standard deviation of 3.4. Find each of the following quantiles, correct to 4 significant figures.

a $t_{0.31}$ b $t_{0.85}$ c $t_{0.224}$ d $t_{0.753}$ e $t_{0.1314}$

2 CAS Find the specified values for each of the following normal variables.

a $\mu = 320, \sigma = 245, P(x < a) = 0.324$

b $\mu = 846, \sigma = 29.7, P(x > a) = 0.592$

c $\mu = 27.8, \sigma = 4.6, P(x \le a) = 0.54$

d $\mu = 39.4, \sigma = 12.6, P(x > a) = 0.82$

e $\mu = 104.5, \sigma = 14.92, P(x > a) = 0.415$

3 A random normal variable X has a mean of 74 and a standard deviation of 8.2.

a What is $P(-\infty \le X \le 64)$?

b Find the value a such that $P(64 \le X < a) = 0.6$.

4 CAS For each of the following random normal variables below, find the specified quantile or value.

a $\mu = 3.63, \sigma = 4.95, P(-\infty < X < a) = 0.254\,627$

b $\mu = 862.4, \sigma = 362.6, P(a < X < 2262.036) = 0.015\,721$

c $\mu = 442, \sigma = 272, P(a < X < \infty) = 0.977\,784$

d $\mu = 720, \sigma = 900, P(-\infty < X < a) = 0.571\,424$

e $\mu = 29.28, \sigma = 12.2, P(13.664 < X < a) = 0.857\,011$

9780170250290

f $\mu = 686, \sigma = 156.8, P(618.576 < X < a) = 0.666\ 384$
g $\mu = 325.5, \sigma = 148.8, P(a < X < 859.692) = 0.882\ 811$
h $\mu = 4260, \sigma = 1740, P(a < X < \infty) = 0.059\ 38$

5 CAS For each of the following random normal variables below, find the specified quantile or value using your CAS calculator.
a $\mu = 418, \sigma = 199.5, P(-\infty < X < a) = 0.053\ 699$
b $\mu = 530.4, \sigma = 102, P(a < X < 913.92) = 0.408\ 961$
c $\mu = 168, \sigma = 196, P(a < X < \infty) = 0.070\ 781$
d $\mu = 326.8, \sigma = 114, P(-\infty < X < a) = 0.492\ 022$
e $\mu = 25.5, \sigma = 37.4, P(46.818 < X < a) = 0.283\ 855$
f $\mu = 86.49, \sigma = 34.41, P(160.1274 < X < a) = 0.010\ 944$
g $\mu = 73.5, \sigma = 20.3, P(a < X < 164.85) = 0.178\ 783$
h $\mu = 1988, \sigma = 1420, P(a < X < \infty) = 0.397\ 432$

Reasoning and communication

6 Example 18 Callum got 18 out of 25 for English and 15 out of 20 for Maths Methods. The means were 15 and 13 and the standard deviations were 8 and 5 for English and Maths Methods respectively. Use Z-scores to determine which subject he did better in.

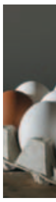
7 Deirdre, who is 17, has an IQ of 110 and a height of 175 cm. The average IQ is 100 and the average height for 17-year-old females is 171 cm with a standard deviation of 12 cm. The standard deviation for IQ is 15. Which is more unusual: Deidre's height as a 17-year-old, or her IQ?

8 Two basketball players have average game totals of 25 points and 29 points respectively. The first has a standard deviation of 9 points and the second a standard deviation of 5 points. Is the first or second player more likely to score more than 37 points in a particular game?

9 Ten-year-old girls can do an average of 11.2 push-ups with a standard deviation of 5.4 push-ups. Which is more unusual for 10-year-old girls: being able to do only 2 push-ups or being able to do 17?

10 The mean and standard deviation of a spelling test were 45 and 13.7. What should the pass mark be set at to ensure that 75% of the people who sat the test passed?

8.09 USING THE NORMAL DISTRIBUTION



The normal distribution is a very good approximation to many situations, particularly for large populations. Real measurements that cluster symmetrically about a mean, with the number of measurements diminishing as the distance from the mean increases, are often approximately normal distributions. The graphs of such measurements are often 'bell-shaped'. Examples include heights and weights of people, other measurements of living things, capacities of manufactured containers and masses of products packed by machine. Most errors are normally distributed, so continuous measurements are usually assumed to be normally distributed.

Example 20

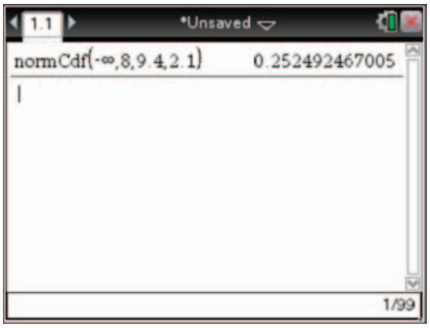
A study of digestion rates found that normal meals were fully digested in an average time of 9.4 hours, with a standard deviation of 2.1 hours. Use this information to find the percentage of people who will digest a normal meal in 8 hours or less.

Shutterstock.com/Monkey Business Images

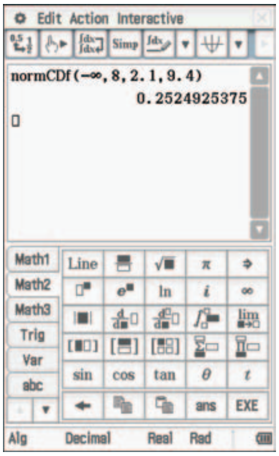
Solution

Write the required proportion for the normal cumulative distribution function. $P(-\infty < X \leq 8)$

TI-Nspire CAS



ClassPad



Write the answer. About 25% of people will have fully digested the meal in 8 hours or less.

When someone says they are 172 cm tall, there is an implied assumption that the measurement is taken to the nearest centimetre. This means that their height is actually between 171.5 cm and 172.5 cm. In the same way, the question 'What is the probability that a Year 12 boy will weigh 76 kg?' is actually asking for the probability that his weight is between 75.5 kg and 76.5 kg.

Example 21

In Australia, Year 12 girls have an average height of 171.9 cm, with a standard deviation of 9.3 cm. What is the probability that a Year 12 girl selected at random will have a height of 180 cm?

Solution

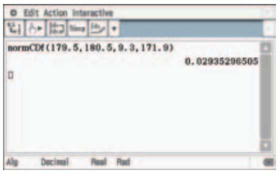
Write the required probability.

$P(179.5 \le X < 170.5)$

TI-Nspire CAS

ClassPad

Use the Rotate option to view the screen lengthways.



Write the answer.

The probability that a Year 12 girl has a height of 180 cm is about 0.0294.

Example 22

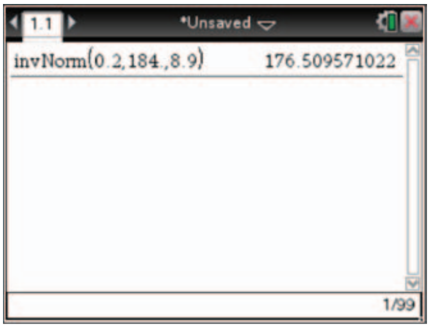
The average height of Year 12 boys is 184.0 cm with a standard deviation of 8.9 cm. Estimate the height of a boy shorter than 80% of all Year 12 boys.

Solution

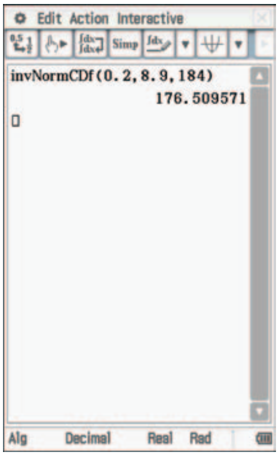
Change to a quantile.

$P(-\infty < X < a) = 0.2$

TI-Nspire CAS

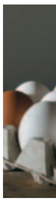


ClassPad



Write the answer.

A Year 12 boy who is 176.5 cm tall will be shorter than 80% of all Year 12 boys.



EXERCISE 8.09 Using the normal distribution

Reasoning and communication

WS Normal distribution – Worded problems 1

WS Normal distribution – Worded problems 2

1 Example 20 The 'life' of young women's fashions measures how long a store can expect a fashion to sell in sufficient quantities to be worth keeping in stock. For some types of fashions it is about 2 months (60 days) but with a wide variation, reflected in the standard deviation of 3 weeks (actually 19 days). What percentage of these fashions would be expected to:
   a have a life of over 3 months (90 days)?
   b have a life of less than a month?
   c be a '9-day-wonder' (less than 10 days)?

2 The average rainfall in a northern Queensland town in January is 268.5 mm, with a standard deviation of 84.5. What is the probability that the rainfall for a particular January is:
   a above 220 mm? b less than 190 mm?

3 The average earnings of plumbers are \$36/h, with a standard deviation of \$7.
   a From a group of 20 plumbers, how many would you expect to be earning over \$43/h?
   b From 50 plumbers, how many would you expect to be earning less than \$38/h?

4 Example 21 Terri is an office worker who starts work at 9:00 a.m. She has found that it takes her an average of 40 minutes to get to work, with a standard deviation of 3 minutes. If she leaves home at 8:15 to allow some extra time, what is the probability that she gets to work at:
   a 9:00 a.m. b 8:59 a.m. c 9:01 a.m. d 8:58 a.m. e 9:02 a.m.
   f What is the probability that she arrives early?
   g What is the probability that she is late?

5 A large factory has 3000 fluorescent tubes. The tubes have an average life of 4500 hours with a standard deviation of 400 hours. What is the probability that a particular tube has a life of:
   a 4500 hours b 4000 hours c 4800 hours
   d 5000 hours e 4100 hours?

6 Avalon Airport has an average maximum temperature in August of 15.5° with a standard deviation of 2.6°. What is the probability that 15 August this year will have a maximum of 15.5°? What is the probability that the maximum will be about 15°C?

7 Example 22 A car battery manufacturer found that their batteries have an average life of 26 months, with a standard deviation of 2.5 months. For what period could they guarantee the batteries if they want no more than 3% to fail within the guarantee period?

8 40 playing cards are dealt at random from a well-shuffled 'shoe' containing 4 ordinary packs in an ESP test. The people being tested have to guess the correct suit for each card as it is dealt. The psychologist doing the test has calculated that they will get an average of 10 correct, with a standard deviation of 2.7. Even if someone does have 'second-sight', she does not expect them to get the suit right every time. She has decided to use a 5% probability level as her threshold to find people who might have 'second-sight'.
   a What number would someone have to get right in order to have less than a 5% chance of doing it by chance?
   b When she did the test on 100 people, she found 7 who exceeded the threshold. Is this sufficient evidence to say that some people do have 'second-sight'?

9780170250290

9 The length of bolts produced by a machine is supposed to be 40 mm. The bolts produced actually have an average length of 40 mm with a standard deviation of 0.6 mm. The manager does not want to reject more than 2% as being too long or short. What is the range of length for this level of acceptability?

10 The towns of Weldon and Betterdon have average annual rainfalls of 1048 mm and 839 mm respectively. The standard deviation for Weldon is 255 mm and for Betterdon it is 122 mm. In which town is the annual rainfall more likely to fall below 500 mm?

11 CAS IQ is normally distributed with a mean of 100 and a standard deviation of 15. Students entering university from school are thought to have an IQ of at least 105.

a What is the probability of a person having an IQ over 120?

b Assuming university students do have an IQ over 105, what is the probability that a university student has an IQ of 120 or more? Hint: Use conditional probability.

12 CAS Potatoes are harvested when the tops brown off. Potatoes between 3 cm and 5 cm in diameter are packed as 'chat' potatoes, those smaller than 3 cm are discarded and others are packed as normal potatoes. The average diameter of harvested potatoes is 7 cm with a standard deviation of 2 cm.

a What is the probability that a harvested potato will be a 'chat' potato?

b What is the probability that a packed potato will be a normal potato?

Alamy/Arco Images/Huetter, C.

8 CHAPTER SUMMARY
CONTINUOUS RANDOM SAMPLES AND THE NORMAL DISTRIBUTION

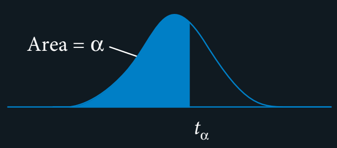
- The **closed interval** $a \le x \le b$ includes the end values a and b and is written as $[a, b]$.
- The **open interval** $a < x < b$ excludes the end values a and b and is written as (a, b).
- The **semi-closed interval** $a \le x < b$ includes a and excludes b and is written as $[a, b)$.
- A **random variable** has a numerical value that depends on the outcome of a chance experiment. It is usually denoted by a capital letter. The corresponding lower-case letter denotes specific values of the variable.
- The **range** (values) of a **discrete** random variable is a finite set, the counting numbers or an equivalent set.
- The range of a **continuous** random variable is an interval of real numbers. The interval can be $(-\infty, \infty)$, the entire set of real numbers.
- The **probability density function** (**pdf**) of a continuous random variable X defined on the interval (a, b) is a function $p(x)$ such that $p(x) \ge 0$ for $x \in (a, b)$ and $\int_a^b p(x)dx = 1$.
  $P(c \le X \le d) = \int_c^d p(x)dx$ for all intervals (c, d) on which X is defined.
- $P(c \le X \le d)$, $P(c < X \le d)$, $P(c \le X < d)$ and $P(c < X < d)$ are all the same because the area of the interval is the same for open and closed intervals.
- The **cumulative distribution function** (**cdf**) for the probability density function $f(x)$ of a continuous random variable X defined on the interval (a, b) is given by $F(x) = \int_a^x p(x)dx$.
- The **expected value** of a continuous random variable X defined on the interval $[a, b]$ is calculated from the probability density function $p(x)$ as $\mu = E(X) = \int_a^b x \cdot p(x)dx$.
- The variance of a continuous random variable X is calculated using $Var(X) = \int_a^b p(x)(x-\mu)^2 dx$, where X is defined on $[a, b]$, $p(x)$ is the probability density function and $\mu = E(X)$.
- The **standard deviation** of a continuous random variable X is given by $SD(X) = \sigma = \sqrt{Var(X)}$.
- A **uniform continuous probability variable** (**rectangular distribution**) is one whose probability density function has a constant value on the domain of X. If X is defined for the domain $[a, b]$, then

$$p(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{otherwise (i.e., } x \notin [a,b]) \end{cases}$$

  The expected value is given by $E(X) = \frac{1}{2}(a+b)$, the variance is $Var(X) = \dfrac{(b-a)^2}{12}$ and the standard deviation is $\sigma = \dfrac{(b-a)}{2\sqrt{3}}$
- A **triangular continuous random variable** is one whose probability density function $p(x)$ has a graph in the shape of a triangle. A **symmetrical triangular distribution** has expected value $E(X) = \frac{1}{2}(a+b)$, variance $Var(X) = \dfrac{(b-a)^2}{24}$ and standard deviation $\sigma = \dfrac{(b-a)}{2\sqrt{6}}$.

9780170250290

- A **linear change of scale and origin** of a random variable X gives a new random variable $Y = aX + b$, where a and b are constants. The statistics are related by: $E(Y) = aE(X) + b$, $Var(Y) = a^2 Var(X)$ and $SD(Y) = aSD(X)$.
- The **standard normal distribution** has the probability density function $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$.
  The mean and standard deviation of the standard normal distribution are $\mu = 0$ and $\sigma = 1$.
- A **normal distribution** is a linear transformation of the standard normal distribution with a probability density function of the form
  $$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
  where μ and σ are the mean and standard deviation of the random variable.
- The standard normal score z **of a random normal value x is obtained by** $Z = \frac{x-\mu}{\sigma}$, where μ and σ are respectively the mean and standard deviation of X. The standard normal score is called the Z-score.
- The **standard normal score** may be used to compare values in different normal distributions.
- A **quantile** t_α for a continuous random variable X is the value of x such that $P(X < t_\alpha) = \alpha$, where $0 < \alpha < 1$. It is the value that divides the distribution in the ratio $\alpha : 1 - \alpha$.



- The **median** is the value such that $\alpha = 0.5$: $P(X < m) = 0.5$.
- The **first quartile** and **third quartile** of a continuous random variable are the quantiles for $\alpha = 0.25$ and $\alpha = 0.75$ respectively.

CHAPTER REVIEW

CONTINUOUS RANDOM SAMPLES AND THE NORMAL DISTRIBUTION

1 Example 3 Which of the following could be probability density functions on the interval [1, 5]?

I $f(x) = 0.4x$

II $f(x) = \dfrac{1}{x\ln(5)}$

III $f(x) = 0.56x - 0.08x^2 - 0.47$

A I B II C III D II and III E I and III

2 Example 5 The cumulative distribution function of the continuous random variable X on the interval [–1, 1] is $\text{cdf}(x) = 0.5x^3$. The probability density function is:

A $1.5x^2$ B $\dfrac{x^4}{8}$ C $0.15x^2$ D $\dfrac{x^3}{6}$ E $\dfrac{x^2}{6}$

3 Example 9 The expected value of a uniform probability density function on the interval [4, 44] is

A $0.0208\overline{3}$ B $0.0\overline{227}$ C 0.025 D 22 E 24

4 Example 12 The variance of a continuous random variable defined on the interval [16, 40] with symmetrical triangular distribution is:

A $2\sqrt{6}$ B 27 C 48 D $4\sqrt{3}$ E 24

5 Example 17 What is the area under the standard normal distribution for $0 \le Z \le 2.14$?

A 0.4780 B 0.4793 C 0.4823 D 0.4838 E 0.4840

6 Example 18 What is the standard normal score of a score of $x = 28$ for a random normal variable with mean 24 and standard deviation 5?

A 0.45 B 0.8 C 1.17 D 1.25 E 1.4

7 Example 17 X is a random normal variable with a mean of 21 and a standard deviation of 5.3. Find the probability that a random value of X lies between 19 and 22.

A 0.0448 B 0.2219 C 0.3773 D 0.5660 E 0.7547

8 Example 19 A random normal variable G has a mean of 50 and a standard deviation of 12. What is the value of g such that 30% of the scores are below g?

a 32.8 b 38.2 c 39.9 d 43.7 e 48.6

Short answer

9 Example 2 The distances driven by some Year 12 students with learner's licences during the last fortnight were as follows.

217, 152, 127, 101, 110, 330, 301, 308, 127, 161, 136, 199, 136, 138, 158, 106, 250, 198, 75, 102, 320, 58, 111, 133, 127, 113, 147, 373, 108, 368, 207, 144, 176, 132, 150, 117, 237, 125, 224, 262, 119, 217, 283, 113, 159, 175, 145, 244, 253, 172, 124, 109, 290, 242, 144, 94, 125, 188, 225, 165, 357, 131, 289, 293, 229, 194, 137, 99, 179, 147, 189, 116, 144, 138, 273, 166, 150, 216, 119, 187, 136, 225, 74, 108, 296, 144, 121, 173, 221, 147

Draw a histogram of the distances with 50 km class widths and use your graph to find the probability that a randomly selected Year 12 learner driver will drive between 180 and 220 km inclusive in a fortnight.

10 Example 4 Make a probability density function using $x^3 - 2x^2 + 2$ on the interval [2, 5].

11 Examples 6, 7 An unreinforced concrete path can crack anywhere along its length. An unreinforced clothesline path is 6 m long. Construct a probability density function for the distance of the first crack from the beginning of the path and hence find the probability that the first crack is 2.3 m from the beginning of the path.

12 Example 10 Find the expected value of a continuous random variable with a triangular probability density function defined on the interval [3, 27] with a maximum value at 18.

13 Examples 13, 14 A continuous random variable X is transformed to the random variable Y according to the equation $Y = 2X + 3$. The mean and standard deviation of X are 27.8 and 5.6 respectively. What are the mean and standard deviation of Y?

14 Example 15 What is the probability density function of a normal distribution with $\mu = 76$ and $\sigma = 5.2$?

15 Example 18 Danielle's class got an average of 18.8 on an English test with a standard deviation of 5.4. The same group scored an average of 22.3 on a Maths Methods test with a standard deviation of 3.6. Danielle scored 27 on both tests. In which test did she do better?

16 Example 17 M is a standard normal variable. Calculate each of the following.

a $P(M > -0.7)$

b $P(0.2 \leq M \leq 2.4)$

17 Example 19 The mean odometer reading of cars pulled up for safety checks was found to be 124 000 km, with a standard deviation of 38 000 km. Assuming the odometer readings were normally distributed, what was the first quartile?

Application

18 Customers in a material shop need to wait until an assistant is available to measure and cut cloth from a roll. The waiting time can be anything from 0 to 15 minutes, with equal chances of any time in between. What is the probability of waiting between 5 and 8 minutes?

19 200 students were asked to estimate the temperature outside. The estimates varied from 23° to 35°. Use a symmetrical triangular distribution to find the probability that a randomly selected student made an estimate of 30°.

20 Students learning to use a pottery wheel take an average time of 25 minutes to make a simple pot. 30% of such students complete their pots within 20 minutes. Assuming that the times are normally distributed, what is the probability of a student taking longer than 28 minutes?

Qz
Practice quiz