Chapter 5 Additional Worksheets

The Normal Distribution

These additional worksheets mirror Chapter 5 (The Normal Distribution) of the AQA Use of Mathematics textbook, exercise by exercise. The numbers, contexts and notation match the same skills as the original in the same order, so you can use this worksheet for extra practice or as a parallel assessment.

Reference: features of a normal distribution (used throughout this worksheet)

A normal distribution is symmetrical about the mean and bell-shaped.
About $\frac{2}{3}$ or 67% of the data lie within 1 standard deviation (s.d.) of the mean.
About 95% of the data lie within 2 s.d. of the mean.
Virtually all the data lie within 3 s.d. of the mean.

The standard normal distribution $N(0, 1)$ has mean 0 and s.d. 1. The area under the curve is 1.
$\Phi(z)$ denotes $P(Z < z)$ for the standard normal distribution.
$\Phi(-z) = 1 - \Phi(z)$

To standardise a value x from $N(\mu, \sigma^2)$: $\quad z = \dfrac{x - \mu}{\sigma}$

5.1 Features of a normal distribution

Whether you are designing a kitchen, a car, or a packaging line, you will have to pay great attention to 'typical' dimensions such as length, weight and time.

You will not be surprised that histograms of many of these sorts of measurements have a bell-shaped outline with most measurements in the centre of the distribution. Many such distributions are said to be 'normally distributed' and are good at modelling not just human and animal measurements, but also variables such as reaction times, manufacturing errors, weights of mechanically-filled packets, and battery life spans.

Activity 1
A workshop produces metal rods for a packaging system. The length of a rod can be modelled by a normal distribution with mean 80 mm and standard deviation 5 mm. The packaging tube used to ship the rods can hold rods up to 90 mm long.
Out of 100 000 rods produced, how many would you expect to have

- **a)** at least 10 mm of spare room in the tube
- **b)** too great a length for the tube?

You should think of the normal distribution as a 'mathematical model' for the actual distribution. This is illustrated in Example 1, where a particular normal distribution models the data well in some ways, but less well in others.

Example 1 (worked, for reference)

A fitness-tracker company recorded the daily step counts of 400 million users on one day.

a) What percentage of users took over 8000 steps?

b) The step counts can be modelled by a normal distribution with mean 7000 and standard deviation 1000. What proportion of users does the model predict to have taken over 8000 steps?

Solution.

a) Suppose the histogram shows that roughly 63 million of the 400 million users took over 8000 steps. As a percentage, $\frac{63}{400} \times 100 \approx 15.75\%$.

b) Over 8000 steps is everything more than 1 s.d. above the mean. For a normal distribution about $\frac{2}{3}$ of the data are within 1 s.d. of the mean, so the other $\frac{1}{3}$ is equally split between values above 8000 and values below 6000. Over 8000 steps is therefore approximately $\frac{1}{6}$ or 16.7% of the distribution – reasonably close to the figure from the histogram.

Key point
For a normal distribution of given mean and standard deviation, think in terms of numbers of standard deviations above (or below) the mean.

Exercise 5A

1. A distribution of weights has mean 65 kg and standard deviation 7 kg. Complete this table of values.

Weight (kg)	72	51	79		
No. of s.d. above mean				-0.5	1.5

2. State whether the distributions described below appear to form a normal distribution. If a distribution is not normal, give the reason.

   a) A histogram of the ages (in completed years) at which residents of a small town retired last year: roughly symmetric, single-peaked, bell-shaped, centred around age 65.

   b) A histogram of the number of text messages sent per day by a sample of teenagers: heavily skewed, with a tall peak near 0 and a long tail stretching out to very high values.

3. Estimate the mean and standard deviation of each of the normal distributions described below (each is given as a smooth, symmetric bell-shaped curve over the stated range of x).

   a) A curve centred and peaking at $x = 12$, with the bulk of the area lying between $x = 9$ and $x = 15$.

   b) A curve centred and peaking at $x = 50$, with the bulk of the area lying between $x = 38$ and $x = 62$.

   c) A curve centred and peaking at $x = 100$, with the bulk of the area lying between $x = 40$ and $x = 160$.

   d) A curve centred and peaking at $x = 1200$, with the bulk of the area lying between $x = 1140$ and $x = 1260$.

4. Adult female reaction times to a visual stimulus are normally distributed with mean 0.25 seconds and standard deviation 0.03 seconds. Adult male reaction times are normally distributed with mean 0.22 seconds and standard deviation 0.025 seconds.

   a) Comment on the differences between the distributions of female and male reaction times.

   b) Explain why the proportion of women with a reaction time slower than 0.31 seconds is the same as the proportion of men with a reaction time slower than 0.27 seconds. What is this proportion?

5. The daily returns on a particular investment fund have a mean of 0.04% and a standard deviation of 0.20%. *These figures are based upon historic performance over the last 20 years and are not guaranteed for the future.*

   a) Comment on the significance of the value of the mean.

   b) Comment on the value of the standard deviation compared to that of the mean.

   c) Assume that the daily returns are normally distributed.

   i On how many days in a year would you expect the daily return to be less than -0.36%?
   ii On how many days in a year would you expect the daily return to be greater than 0.64%?

Normal distributions play a major role in modelling customer demand. This is vital for many businesses so that the right amount of stock is kept in store.

6. **a)** Why do most businesses keep a stock of products for which they do not yet have orders?

   b) What is the disadvantage to a business of choosing to keep a larger than necessary amount of stock?

c) An online bookshop models its sales per year by a normal distribution with mean 500 orders per day and standard deviation 25 orders per day. According to this model, how likely is it that on a given day sales will be less than 450 orders?

7. The weights of packets of pasta are normally distributed with mean 500 g and standard deviation 4 g. A random sample of 300 packets is selected and each packet is weighed.

   a) How many packets would you expect to have a weight outside the range from 495 g to 505 g?

   b) How many packets would you expect to have a weight in the range from 497.5 g to 502.5 g?

   c) How many packets would you expect to have a weight in the range from 497.5 g to 500 g?

8. The number of calls per minute to a customer-service hotline at peak time can be modelled by a normal distribution with mean 25 and standard deviation 6.

   a) How likely is it that the number of calls per minute will be more than 40?

   b) How likely is it that the number of calls per minute will be less than 15?

   c) How likely is it that the number of calls per minute will be between 15 and 35?

5.2 Calculating probabilities

Example 2 (worked, for reference)

Find the proportion of a standard normal distribution which lies between $z = -0.3$ and $z = 1.6$.

Solution. You need $\Phi(1.6) - \Phi(-0.3)$.
$\Phi(1.6) = 0.94520$
$\Phi(-0.3) = 1 - \Phi(0.3) = 1 - 0.61791 = 0.38209$
$\Phi(1.6) - \Phi(-0.3) = 0.94520 - 0.38209 = 0.56311$
So 56.311% of the distribution lies between $z = -0.3$ and $z = 1.6$.

Activity 2
State the approximate values of $\Phi(-2)$, $\Phi(-1)$, $\Phi(0)$, $\Phi(1)$ and $\Phi(2)$. Think about the symmetry of the graph.

Activity 3
What are the following probabilities?

a) $P(z = 1)$
b) $P(-0.4 \leqslant z \leqslant 1.7)$
c) The probability of an impossible event.
d) The probability of a certain event.

Example 3 (worked, for reference)

For the distribution $N(0, 1)$:

a) Find **i** $P(z < 1.3)$ **ii** $P(-1 < z < 1.3)$ **iii** $P(z > -1.2)$
b) Find the number a such that $P(z < a) = 0.81594$
c) Find a symmetric interval about 0 such that the probability of a randomly chosen member of the distribution lying in this interval is 0.92

Solution.
a) i) $\Phi(1.3) = 0.90320$.
ii) $\Phi(1.3) - \Phi(-1) = 0.90320 - (1 - 0.84134) = 0.74454$.
iii) $1 - \Phi(-1.2) = \Phi(1.2) = 0.88493$.
b) $\Phi(a) = 0.81594$ so $a = 0.9$.
c) $P(-b < z < b) = 0.92$ means $\Phi(b) = 0.96$, so $b \approx 1.75$. The interval is $(-1.75, 1.75)$.

Exercise 5B

1. **a)** Find **i** $P(-1 < z < 1)$ **ii** $P(-2 < z < 2)$ **iii** $P(-3 < z < 3)$

   b) Compare your answers with 95%, $\frac{2}{3}$ and 1.

2. Find

   a) $P(z > 1.4)$

   b) $P(z < -1.6)$

   c) $P(0.8 < z < 2.2)$

5.3 Standardising a normal variable

Example 4 (worked, for reference)

Suppose that the heights of 9-year-old girls are normally distributed with a mean of 133.5 cm and standard deviation of 11.2 cm.

a) What is the probability that a 9-year-old girl will have a height of less than 140 cm?

b) How many of a random sample of 800 9-year-old girls would you expect to be less than 140 cm tall?

Solution.

a) $z = \dfrac{140 - 133.5}{11.2} = 0.580$, so $P(h < 140) = \Phi(0.580) \approx 0.719$.

b) Expected number $\approx 800 \times 0.719 = 575$.

Key point

The notation for a normal distribution with mean μ and s.d. σ is $N(\mu, \sigma^2)$. The second number is the square of the s.d. and is called the **variance**.

If a value x is taken from $N(\mu, \sigma^2)$ then the z-value is given by $z = \dfrac{x - \mu}{\sigma}$. This process is called **standardising** the variable.

Example 5 (worked, for reference)

The weights of the contents of jars of jam are normally distributed with mean 340 g and s.d. 5 g.

a) Find the z-values of **i** 332 g **ii** 347 g.

b) Find the probability that the contents of a jar of jam have a weight between 332 g and 347 g.

Solution.

a) i) $\dfrac{332 - 340}{5} = -1.6$ ii) $\dfrac{347 - 340}{5} = 1.4$

b) $P(332 < \text{weight} < 347) = \Phi(1.4) - \Phi(-1.6) = 0.91924 - (1 - 0.94520) = 0.86444 \approx 0.86$.

Example 6 (worked, for reference)

The length of time a particular curry house takes to deliver food on a Saturday evening has distribution $N(35, 81)$. At what time should you order if you want a 92% chance of it arriving by 8 pm?

Solution. $\Phi(1.405) \approx 0.92$, so there's a 92% chance of it taking less than 1.405 standard deviations above the mean. $35 + 1.405 \times 9 \approx 47.6$ minutes. You should order at 7:12 pm.

Example 7 (worked, for reference)

A reading speed, in words per minute, of a randomly selected adult has distribution $N(250, 40^2)$.

a) Find the probability that a person reads at a speed of over 350 wpm.

b) Find the reading speed below which 90% of the population lie.

Solution.

a) $350 = 250 + 2.5 \times 40$, so $P(z > 2.5) = 1 - \Phi(2.5) = 0.00621$, or 0.6%.

b) $\Phi(1.28) \approx 0.9$, so the required speed is $250 + 1.28 \times 40 = 301.2$ wpm.

Activity 4

A condition of membership of a particular high-achievers' society is that you should score over 700 on a certain standardised test. The society's website states that it is a society for the top 4% of test-takers. Use the fact that the test is designed to have the distribution $N(500, 110^2)$ to check if scoring over 700 is the same as being in the top 4% of test-takers.

Exercise 5C

1. State the value of the standard deviation for each of these distributions.
   a) $N(8, 6^2)$
   b) $N(8, 49)$
   c) $N(21, 60)$

2. Calculate z-values for each of these.
   a) A value of 11 in a normal distribution with mean 7 and standard deviation 4.
   b) A value of 5 in a normal distribution with mean 9 and standard deviation 12.
   c) A value of -2 in $N(3, 5^2)$.
   d) A value of 7 in $N(5, 16)$.
   e) A value of 2 in $N(4, 9)$.
   f) A value of 6 in $N(8, 8)$.

3. For a standard normal distribution, find these probabilities.
   a) $P(z < 1.32)$
   b) $P(z > 1.87)$
   c) $P(z < -2)$
   d) $P(z > -4)$
   e) $P(0.5 < z < 2.1)$
   f) $P(-2 < z < 1)$
   g) $P(-2.18 < z < -1.33)$

4. **a)** For a standard normal distribution, explain, using a diagram, why $P(z < -a) = P(z > a)$, for any positive number a.
   b) Hence explain why $P(-a < z < a) = 2P(z < a) - 1$.

5. For a standard normal distribution, find the value of the number a for each of these.
   a) $P(z < a) = 0.722$
   b) $P(z > a) = 0.02$
   c) $P(z < a) = 0.15$
   d) $P(-a < z < a) = 0.90$
   e) $P(-a < z < a) = 0.50$

6. Assume that typing speeds, R (words per minute), are normally distributed with mean 72 and variance 12. Find numbers $72 - c$ and $72 + c$ for each of these probabilities.
   a) $P(72 - c < R < 72 + c) = 0.34$
   b) $P(72 - c < R < 72 + c) = 0.754$

   Hint: the answer to a) is the z-value of $72 + c$. First find the value of a such that $P(-a < z < a) = 0.34$.

7. Even with measurements of such quantities as reaction times and lung capacities, the distributions

for some populations will not be normally distributed. Suggest possible reasons why a distribution might have:

a) a sharp spike at one particular value with very little spread either side

b) two separate peaks rather than one.

Consolidation Exercise 5

1. The weights of kittens at birth follow a normal distribution with a mean of 1.2 kg and a standard deviation of 0.18 kg. New-born kittens are classified as small, medium or large, according to their weight. Kittens with a weight of more than 1.5 kg are classified as large.

   The probability of a new-born kitten being classified as small is 0.137.

   a) Calculate the probability that a new-born kitten is classified as large.

   b) Calculate the maximum possible weight of a kitten that is classified as small.

2. Body temperatures of dogs may be assumed to be normally distributed with a mean of 38.5 °C and a standard deviation of 0.4 °C.

   Calculate the probability that a dog, chosen at random, has a body temperature greater than 40.0 °C.

3. The heights of girls aged 9 years may be assumed to be normally distributed with a mean of 133.5 cm and a standard deviation of 11.2 cm. Calculate the probability that a girl, chosen at random, has a height

   a) greater than 136 cm

   b) less than 138 cm

   c) less than 128 cm

   d) between 132 and 135 cm.

4. Each evening, Idris completes a sudoku puzzle in his local newspaper. His completion times, X minutes, can be modelled by a normal random variable with a mean of 18 and a standard deviation of 6.

   Determine

   a) $P(X < 25)$

   b) $P(X > 15)$

5. Hana is employed by a telecoms company to install routers in new homes on a housing estate. Her time, X minutes, to install a router may be assumed to be normally distributed with a mean of 35 and a standard deviation of 15.

   Determine

   a) $P(X < 45)$

   b) $P(20 < X < 45)$

   c) the time, k minutes, such that $P(X < k) = 0.9$

6. Assume that the ankle circumference of the adult male population of the UK is normally distributed with a mean of 225 mm and a standard deviation of 18 mm.

   a) A manufacturer designs an ankle support for the UK market. It is designed to fit ankle circumferences up to 250 mm. What percentage of the UK adult male population is catered for by this support?

   b) A support designed for another market will fit ankles with circumferences from 195 mm to 245 mm. Find the percentage of the UK adult male population that this support will fit.

7. The systolic blood pressure of a randomly-selected adult can be assumed to be normally distributed

with mean 122 and standard deviation 15.

Find the probability that

a) **i** an adult has a systolic blood pressure over 130 **ii** an adult has a systolic blood pressure between 110 and 125

b) Find the blood pressure below which 90% of the population lie.

Review

After working through this chapter you should:

- know the main features of a normal distribution
- be familiar with the notation $N(\mu, \sigma^2)$
- be able to find z-values for any normal distribution
- be able to use either tables or a calculator to find the probability of an event that is normally distributed.

The central limit theorem

The normal distribution is the single most important idea in statistics. Some aspects of the result that underlies it – the central limit theorem – can be investigated very simply.

Take some coins and throw them repeatedly. Count the number of heads each time and then draw a histogram of your results. Your histogram is likely to show some, but not all, of the features of a normal distribution.

Why the normal distribution should appear in so many different contexts was explained when the central limit theorem was discovered. You can research the central limit theorem, and the mathematicians who contributed to it, on the internet.