8 THE NORMAL DISTRIBUTION

Objectives

After studying this chapter you should

- appreciate the wide variety of circumstances in which the normal distribution can be used;
- be able to use tables of the normal distribution to solve problems;
- be able to use the normal distribution as an approximation to other distributions in appropriate circumstances.

Statistical tables are available in many books and can also be found online.

You should note that the accuracy of your solutions will depend on the tables (or graphic calculator) you are using.

8.0 Introduction

The tallest accurately recorded human being was *Robert Wadlow* in the USA. On his death at the age of 22 he was 272 cm (8 feet 11.1 inches) tall. If you were an architect and you had to design doorways in a building you would clearly not make them all 9 feet high - most ceilings are lower than this!

What height should the ceilings be?

In 1980 the Government commissioned a survey, carried out on 10 000 adults in Great Britain. They found that the average height was 167.3 cm with SD (standard deviation) 9.1 cm. You cannot make a door size that everyone can fit through but what height of door would 95% of people get through without stooping? This chapter should help you find the answer.

Activity 1   Data collection

There are many sets of data you could collect from people in your group, such as heights, weights, length of time breath can be held, etc. However, you will need about 100 results to do this activity properly so here are a few suggestions where large quantities of data can be collected quickly.

1. **Lengths of leaves**
   Evergreen bushes such as laurel are useful - though make sure all the leaves are from the same year's growth.

2. **Weights of crisp packets**
   Borrow a box of crisps from a canteen and weigh each packet accurately on a balance such as any Science laboratory would have.
3. **Pieces of string**
   Look at 10 cm on a ruler and then take a ball of string and try to cut 100 lengths of 10 cm by guessing. Measure the lengths of all the pieces in mm.
4. **Weights of apples**
   If anyone has apple trees in their garden they are bound to have large quantities in the autumn.
5. **Size of pebbles on a beach**
   Geographers often look at these to study the movement of beaches. Use a pair of calipers then measure on a ruler.
6. **Game of bowls**
   Make a line with a piece of rope on the grass about 20 metres away. Let everyone have several goes at trying to land a tennis ball on the line. Measure how far each ball is from the line.

Try at least two of these activities. You will need about 100 results in all. To look at the data it would help to have a data handling package on a computer.

8.1 Looking at your data

The data shown on the opposite page gives the length from top to tail (in millimetres) of a large group of frogs. This has been run through a computer package so you can see some useful facts about the data.

In the computer analysis you will see that most of the frogs are close to the mean value, with fewer at the extremes. This 'bell-shaped' pattern of distribution is typical of data which follows a normal distribution. To obtain a perfectly shaped and symmetrical distribution you would need to measure thousands of frogs.

Does your data follow a 'bell shaped' pattern?

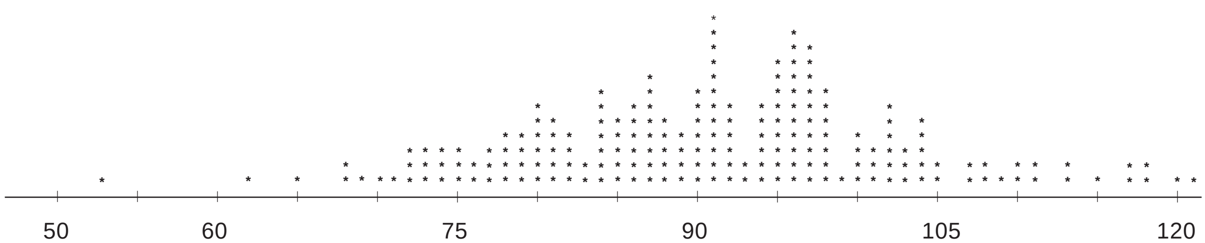
You may notice that median $\approx$ mean $\approx$ mode, as might be expected for a symmetrical distribution. From the analysis of data you also see that the mean is 90.9 mm and the standard deviation is 11.7 mm. Now look at how much of the data is close to the mean, i.e. within one standard deviation of it. From the stem and leaf table you can see that 74 frogs have a length within one standard deviation above the mean and 59 within a SD below the mean.

Altogether, 133 frogs are $+$or $-$ one SD from the mean, which is 66.5%.

Frog Data

The data below show the length from top to tail in millimetres of a large group of frogs.

83	69	97	53	89	95	105	80	76	117	74
91	100	77	110	68	118	87	97	78	100	95
73	103	96	72	71	99	121	81	104	68	89
87	96	87	72	79	102	98	97	88	87	86
103	79	104	105	91	82	102	75	95	90	62
65	97	86	97	111	98	92	74	88	84	80
95	96	92	95	100	90	91	95	75	70	84
80	98	96	94	101	85	113	96	103	98	95
84	84	97	95	108	94	79	81	92	85	87
90	85	82	81	97	79	90	90	94	98	73
91	91	107	102	89	85	98	84	91	90	86
113	86	93	77	100	96	90	97	109	102	84
85	87	97	92	107	102	104	94	93	75	96
91	117	91	87	118	96	89	88	111	120	92
76	94	104	80	77	94	84	78	73	92	81
83	104	91	91	96	88	115	96	74	88	86
80	98	101	95	96	102	78	97	80	87	82
72	78	108	91	91	91	110	86	101	81	97
82	97									



	N	MEAN	MEDIAN	TRMEAN	STDEV	MIN	MAX	Q1	Q3
Frogs	200	90.905	91.000	90.822	11.701	53.000	121.000	83.250	97.000

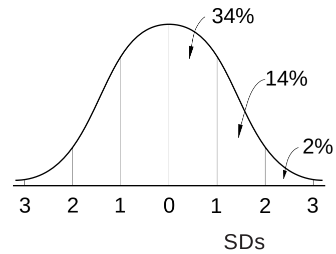
Stem and leaf of frogs

Leaf Unit = 1.0 N = 200

1	5	3
1	5	
2	6	2
6	6	5 8 8 9
17	7	0 1 2 2 2 3 3 3 4 4 4
33	7	5 5 5 6 6 7 7 7 8 8 8 8 9 9 9 9
57	8	0 0 0 0 0 0 1 1 1 1 1 2 2 2 2 3 3 4 4 4 4 4 4 4
85	8	5 5 5 5 5 6 6 6 6 6 6 7 7 7 7 7 7 7 7 8 8 8 8 8 9 9 9 9
(34)	9	0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 3 3 4 4 4 4 4 4
81	9	5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7 8 8 8 8 8 8 8 9
41	10	0 0 0 0 1 1 1 2 2 2 2 2 2 2 3 3 3 4 4 4 4 4
20	10	5 5 7 7 8 8 9
13	11	0 0 1 1 3 3
7	11	5 7 7 8 8
2	12	0 1

Activity 2

Apply the same techniques to your own sets of data (i.e. draw up frequency tables or histograms and calculate means and SDs) and calculate the percentage which lie within one SD of the mean. If the data is **normally** distributed then this should be about 68%.



Similarly, you could look for the amount of data within 2 SDs , 3 SDs, etc. The table below gives approximately the percentages to expect.

Distances from mean in terms of standard deviation in one direction	0 – 1	1 – 2	2 – 3	over 3
Proportion of area in the above range	34 %	14 %	2 %	negligible

Note that very few items of data fall beyond three SDs from the mean.

What is clearly useful is that no matter what size the numbers are, if data are normally distributed, the proportions within so many SDs from the mean are always the same.

Example

IQ test scores, and the results of many other standard tests, are designed to be normally distributed with mean 100 and standard deviation 15.

Therefore statements such as the following can be made:

'68% of all people should achieve an IQ score between 85 and 115.'

'Only 2% of people should have an IQ score less than 70.'

'Only 1 in a 1000 people have an IQ greater than 145.'

Exercise 8A

The survey mentioned in the introduction also showed that the average height of 16-19 year olds was approximately 169 cm with SD 9 cm.

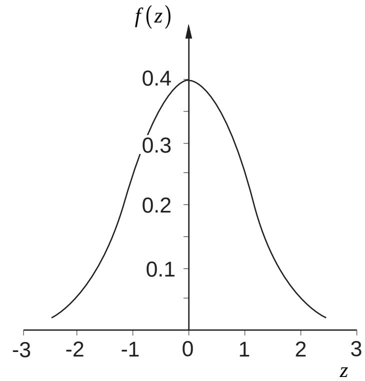
1. Assuming the data follows a normal distribution, find:
   (a) the percentage of sixth formers taller than 187 cm;
   (b) the percentage of sixth formers smaller than 160 cm;
   (c) in a sixth form of 300, the number of students smaller than 151 cm.

(Note these are not truly normal, as the pattern for girls and boys is different.)

8.2 The p.d.f. of the normal

If you could work in only whole numbers of SDs, the number of problems that could be solved would be limited. To calculate the proportions or probabilities of lying within so many SDs of the mean, you need to know the p.d.f. This was first discovered by the famous German mathematician, *Gauss* (1777-1855) and this is why the normal distribution is sometimes called the **Gaussian distribution.**
It is given by the formula

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}.$$



z is called the **standard normal variate** and represents a normal distribution with mean 0 and SD 1. The graph of the function is shown opposite.

Note that the function $f(z)$ has no value for which it is zero, i.e. it is possible, though very unlikely, to have very large or very small values occurring.

In order to find the probabilities of all possible SDs from the mean you would have to integrate the function between the values. This is a tedious task involving integration by parts and to avoid this, tables of the function are commonly used.

z	.00	.01	.02	.03	.04	.05
0.0	.50000	.50399	.50798	.51197	.51595	.5199
0.1	.53983	.54380	.54776	.55172	.55567	.559
0.2	.57926	.58317	.58706	.59095	.59483	.598
0.3	.61791	.62172	.62552	.62930	.63307	.636
0.4	.65542	.65910	.66276	.66640	.67003	.673
0.5	.69146	.69497	.69847	.70194	.70540	.708
0.6	.72575	.72907	.73237	.73565	.73891	.74
0.7	.75804	.76115	.76424	.76730	.77035	.77
0.8	.78814	.79103	.79389	.79673	.79955	
0.9	.81594	.81859	.82121	.82381	.82	
1.0	.84134	.84375	.84614	.84		
1.1	.86433	.86650	.8686			
1.2	.88493	.88686	.8			
1.3	.90320	.90490	.			
1.4	.91924	.92073				
1.5	.93319	.93448				
1.6	.94520	.94630				
1.7	.95543	.95637				
1.8	.96407	.964				
1.9	.97128	.9				

$\Phi(1.0)$

$\Phi(1.5)$

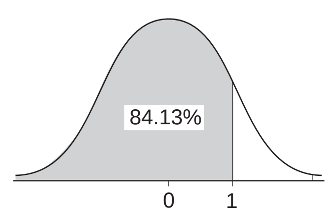
$$\Phi(z) = P(Z < z)$$

$$= \int_{-\infty}^{z} f(z)\, dz.$$

Here Φ has been used to denote the cumulative probability.

For positive z, the function gives you the probability of being less than z SDs above the mean.

For example, $\Phi(1.0) = 0.84313$, therefore 84.13% of the distribution is less than one SD above the mean.



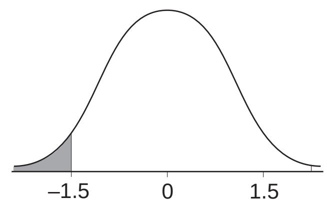
Tables usually give the area to the left of z and only for values above zero. This is because symmetry enables you to calculate all other values.

Example

What is the probability of being less than 1.5 SDs below the mean i.e. $\Phi(-1.5)$?

Solution

From tables,



$$\Phi(+1.5) = 0.93319$$

and by symmetry,

$$\Phi(-1.5) = 1 - 0.93319 = 0.06681$$

i.e. about 6.7%.

A random variable, Z, which has this p.d.f. is denoted by

$$\boxed{Z \sim N(0,1)}$$

showing that it is a normal distribution with mean 0 and standard deviation 1.

This is often referred to as the **standardised** normal distribution.

Example

If $Z \sim N(0,1)$, find

(a) $P(Z > 1.2)$

(b) $P(-2.0 < Z < 2.0)$

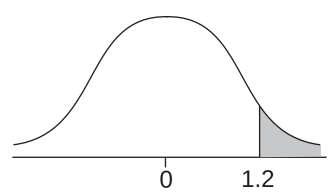
(c) $P(-1.2 < Z < 1.0)$

Solution

(a) $P(Z > 1.2) = 1 - \Phi(1.2)$

$= 1 - 0.88493$ (from tables)

$= 0.11507$



(b) $P(-2.0 < Z < 2.0)$

$= P(Z < 2.0) - P(Z < -2.0)$

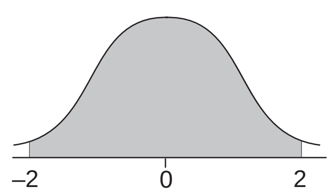
$= \Phi(2.0) - P(Z > 2.0)$

$= \Phi(2.0) - (1 - P(Z < 2.0))$

$= 2\Phi(2.0) - 1$

$= 2 \times 0.97725 - 1$

$= 0.9545$

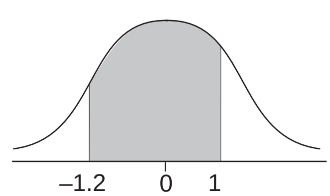


(c) $P(-1.2 < Z < 1.0)$

$= P(Z < 1.0) - P(Z < -1.2)$

$= P(Z < 1.0) - P(Z > 1.2)$

$= \Phi(1.0) - (1 - \Phi(1.2))$



$= 0.84134 - (1 - 0.88493)$

$= 0.72627$

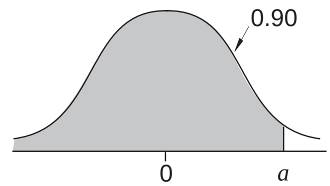
You can also use the tables to find the value of a when $P(Z > a)$ is a given value and $Z \sim N(0,1)$. This is illustrated in the next example.

Example

If $Z \sim N(0,1)$, find a such that

(a) $P(Z < a) = 0.90$

(b) $P(Z > a) = 0.25$



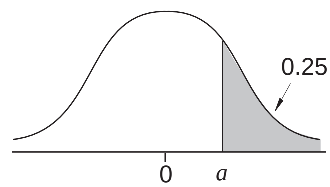
Solution

(a) Here $\Phi(a) = 0.90$, and from the tables

$a \approx 1.28$

(b) Here $\Phi(a) = 1 - 0.25 = 0.75$ and from the tables

$a \approx 0.67$



Exercise 8B

If $Z \sim N(0,1)$, find

1. $P(Z > 0.82)$
2. $P(Z < 0.82)$
3. $P(Z > -0.82)$
4. $P(Z < -0.82)$
5. $P(-0.82 < Z < 0.82)$
6. $P(-1 < Z < 1)$
7. $P(-1 < Z < 1.5)$
8. $P(0 < Z < 2.5)$
9. $P(Z < -1.96)$
10. $P(-1.96 < Z < 1.96)$

8.3 Transformation of normal p.d.f.s

The method needed to transform any normal variable to the standardised variable is illustrated in the example below.

Example

Eggs laid by a particular chicken are known to have lengths normally distributed, with mean 6 cm and standard deviation 1.4 cm. What is the probability of:

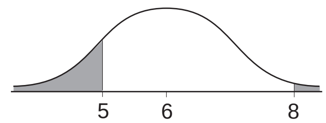
(a) finding an egg bigger than 8 cm in length;

(b) finding an egg smaller than 5 cm in length?

Solution

(a) The number of SDs that 8 is above the mean is given by

$$z = \frac{x-\mu}{\sigma} = \frac{8-6}{1.4} = 1.429 \,,$$



but $\Phi(1.43) = 0.92364$ (from tables)

so $P(X>8) = 1 - 0.92364 = 0.07646$.

(b) $$z = \frac{5-6}{1.4} = -0.7143,$$

but $\Phi(0.7143) = 0.7625$ (from tables using interpolation),

so $P(X<5) = 1 - 0.7625 = 0.2375$.

Note that using interpolation from tables is not necessary for the AEB examination, but it is good practice to use it to improve accuracy.

Note that in order to find the probability you need to establish whether you need the area greater than a half or less than a half. Drawing a diagram will help.

When a variable X follows a **normal distribution**, with mean μ and variance σ^2, this is denoted by

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

So in the last example, $X \sim N\left(6, 1.4^2\right)$.

To use normal tables, the transformation

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

is used. This ensures that Z has mean 0 and standard deviation 1, and the tables are then valid.

Using the UK data on heights in Section 8.0, the z value for Robert Wadlow's height is

$$z = \frac{272 - 167.3}{9.1} \approx 11.5 .$$

So his height is 11.5 SDs above the mean. The most accurate tables show that 6 SDs is only exceeded with a probability of 10^{-10}, so it is extremely unlikely that a taller person will ever appear!

Example

If $X \sim N(4,9)$, find

(a) $P(X > 6)$

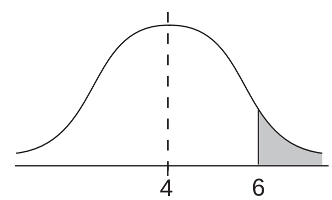
(b) $P(X > 1)$

Solution

Now $Z = \dfrac{X-\mu}{\sigma} = \dfrac{X-4}{3}$,

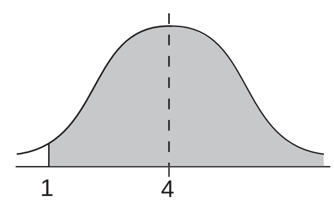
(a) Hence

$$\begin{aligned} P(X>6) &= 1 - P(X<6) \\ &= 1 - \Phi\left(\frac{6-4}{3}\right) \\ &= 1 - \Phi(0.67) \\ &= 1 - 0.74857 \\ &= 0.25143 \end{aligned}$$



(b) $P(X>1) = P(X<7)$ (by symmetry)

$$\begin{aligned} &= \Phi\left(\frac{7-4}{3}\right) \\ &= \Phi(1) \\ &= 0.84134 \end{aligned}$$



Exercise 8C

1. If $X \sim N(200,625)$, find

   (a) $P(X > 250)$ (b) $P(175 < X < 225)$

   (c) $P(X < 275)$

2. If $X \sim N(6,4)$, find

   (a) $P(X > 8)$ (b) $P(4 < X < 8)$

   (c) $P(5 < X < 9)$

3. If $X \sim (-10,36)$, find

   (a) $P(X < 0)$ (b) $P(-12 < X < -8)$

   (c) $P(-15 < X < 0)$

4. Components in a personal stereo are normally distributed with a mean life of 2400 hours with SD 300 hours. It is estimated that the average user listens for about 1000 hours in one year. What is the probability that a component lasts for more than three years.

5. The maximum flow of a river in Africa during the 'rainy season' was recorded over a number of years and found to be distributed $N\left(6300,\ 1900^2\right)$ $\text{m}^3\,\text{s}^{-1}$.

   For the banks to burst a flow of 8700 $\text{m}^3\,\text{s}^{-1}$ is required. What is the probability of this happening in a particular year?

6. IQs are designed to be $N(100,\ 225)$. To join Mensa an IQ of 138 is required. What percentage of the population are eligible to join?

   A psychologist claims that any child with an IQ of 150+ is 'gifted'. How many 'gifted' children would you expect to find in a school of 1800 pupils?

7. Rainfall in a particular area has been found to be $N\left(850,\ 100^2\right)$ mm over the years. What is the probability of rainfall exceeding 1000 mm?

8. In a verbal reasoning test on different ethnic groups, one group was found to have scores distributed $N\left(98.42,\ 15.31^2\right)$. Those with a score less than 80 were deemed to be in need of help. What percentage of the overall group were in need of help?

8.4 More complicated examples

The following examples illustrate some of the many uses and applications of the normal distribution.

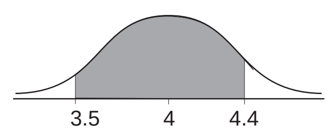
Example

A machine produces bolts which are $N(4,\ 0.09)$, where measurements are in mm. Bolts are measured accurately and any which are smaller than 3.5 mm or bigger than 4.4 mm are rejected. Out of a batch of 500 bolts how many would be acceptable?

Solution

$$P(X < 4.4) = \Phi\left[\frac{(4.4 - 4)}{0.3}\right] \approx \Phi(1.33) = 0.90824$$

$$P(X < 3.5) = \Phi\left[\frac{(3.5 - 4)}{0.3}\right] \approx \Phi(-1.67) = 0.04746.$$



Hence $P(3.5 < X < 4.4) \approx 0.90824 - 0.04746$

$$= 0.86078.$$

The number of acceptable items is therefore

$$0.86078 \times 500 = 430.39 \approx 430 \text{ (to nearest whole number).)}$$

Example

IQ tests are measured on a scale which is $N(100,\ 225)$. A woman wants to form an 'Eggheads Society' which only admits people with the top 1% of IQ scores. What would she have to set as the cut-off point in the test to allow this to happen?

Solution

From tables you need to find *z* such that $\Phi(z) = 0.99$.

This is most easily carried out using a 'percentage points of the normal distribution' table, which gives the values directly.

Now $\Phi^{-1}(0.99) = 2.3263$

which is an alternative way of saying that

$$\Phi(2.3263) = 0.99.$$

(Check this using the usual tables.)

This means that

$$\frac{x-100}{\sqrt{225}} = 2.3263.$$

Hence $x = 100 + 2.3263 \times 15$

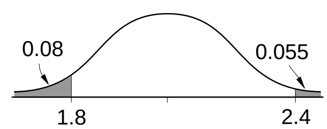
$$= 134.8945 \approx 134.9.$$

Example

A manufacturer does not know the mean and SD of the diameters of ball bearings he is producing. However, a sieving system rejects all bearings larger than 2.4 cm and those under 1.8 cm in diameter. Out of 1000 ball bearings 8% are rejected as too small and 5.5% as too big. What is the mean and standard deviation of the ball bearings produced?

Solution

Assume a normal distribution of



$$\Phi^{-1}(1-0.08) = 1.4;$$

so 1.8 is 1.4 standard deviations below mean.

Also $\Phi^{-1}(1-0.055) = 1.6,$

so 2.4 is 1.6 standard deviations above the mean.

This can be written as two simultaneous equations and solved:

$$\mu + 1.6\sigma = 2.4$$

$$\mu - 1.4\sigma = 1.8.$$

Subtracting,

$$3.0\sigma = 0.6$$

$$\Rightarrow \quad \sigma = 0.2$$

Using the first equation,

$$\mu + (1.6 \times 0.2) = 2.4$$

$$\Rightarrow \quad \mu = 2.4 - (1.6 \times 0.2)$$

$$\Rightarrow \quad \mu = 2.08.$$

So diameters are distributed $N\left(2.08,\ 0.2^2\right)$.

Exercise 8D

1. Bags of sugar are sold as 1 kg. To ensure bags are not sold underweight the machine is set to put a mean weight of 1004 g in each bag. The manufacturer claims that the process works to a standard deviation of 2.4. What proportion of bags are underweight?
2. Parts for a machine are acceptable within the 'tolerance' limits of 20.5 to 20.6 mm. From previous tests it is known that the machine produces parts to $N\left(20.56,\ (0.02)^2\right)$.

   Out of a batch of 1000 parts how many would be expected to be rejected?
3. Buoyancy aids in watersports are tested by adding increasing weights until they sink. A club has two sets of buoyancy aids. One set is two years old, and should support weights according to $N(6.0,\ 0.64)$ kg; the other set is five years old and should support weights of $N(4.5,\ 1.0)$ kg. All the aids are tested and any which are unable to support at least 5 kg are thrown out.

   (a) If there are 24 two-year-old aids, how many are still usable?

   (b) If there are 32 five-year-old aids how many are still usable?
4. Sacks of potatoes are packed by an automatic loader with mean weight 114lb. In a test it was found that 10% of bags were over 116 lb. Use this to find the SD of the process. If the machine is now adjusted to a mean weight of 113 lb, what % are now over 116 lb if the SD remains unaltered?
5. In a soap making process it was found that $6\frac{2}{3}$ % of bars produced weighed less than 90.50 g and 4% weighed more than 100.25 g.

   (a) Find the mean and the SD of the process.

   (b) What % of the bars would you expect to weigh less than 88 g?
6. A light bulb manufacturer finds that 5% of his bulbs last more than 500 hours. An improvement in the process meant that the mean lifetime was increased by 50 hours . In a new test, 20% of bulbs now lasted longer than 500 hours.

   Find the mean and standard deviation of the original process.

8.5 Using the normal as an approximation to other distributions

In earlier chapters you looked at discrete distributions such as the binomial. Let us suppose that the probability of someone buying the *Daily Sin* newspaper in a particular town is 0.4. Consider these problems:

(a) What is the probability that in a row of six houses all six buy the *Sin*?

(b) Of 25 customers who come into a shop what is the probability of 10 or more buying the *Sin*?

(c) Two hundred people live on an estate. What is the probability that 100 or more buy the *Sin*?

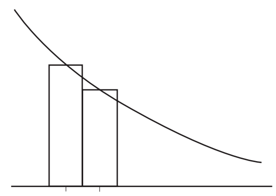
In part (a) you would probably use the binomial distribution and a calculator to find $(0.4)^6$ and in (b) you would probably use tables to save on calculation. However, in part (c) there is a problem. Tables do not go beyond 50; you could use a Poisson approximation, but *p* is not really small and this would still involve enormous calculations. Imagine a probability histogram with 200 columns – it would look almost continuous! You will already know that for *p* approximately half, you get a symmetrical bell shaped graph. In fact you can use the normal distribution as an approximation in such cases.

You know that for a binomial distribution

$$\mu = np = 200 \times 0.4 = 80$$

and $$\sigma^2 = np(1-p) = 200 \times 0.4 \times 0.6 = 48$$

$$\Rightarrow \quad \sigma = 6.93.$$



A slight adjustment needs to be made since the 100 column actually goes from 99.5 to 100.5. To include 100 you need to find $P(x > 99.5)$. This is sometimes called a **continuity correction** factor.

So P(100 or more buy *Sin*) $= 1 - \Phi\left(\dfrac{99.5 - 80}{6.93}\right)$

$$= 1 - \Phi(2.81)$$

$$= 1 - 0.99752 \quad \text{(from tables)}$$

$$= 0.00248.$$

In the same way you can use the normal distribution to approximate for the Poisson.

Example

Customers arrive at a garage at an average rate of 2 per five minute period. What is the probability that less than 15 arrive in a one hour period?

Solution

$\mu = \sigma^2 = 24$ per hour, so $\sigma = 4.9$.

Hence $P(\text{less than 15 in an hour}) = \Phi\left(\dfrac{14.5 - 24}{4.9}\right)$

$$\approx \Phi(-1.94)$$

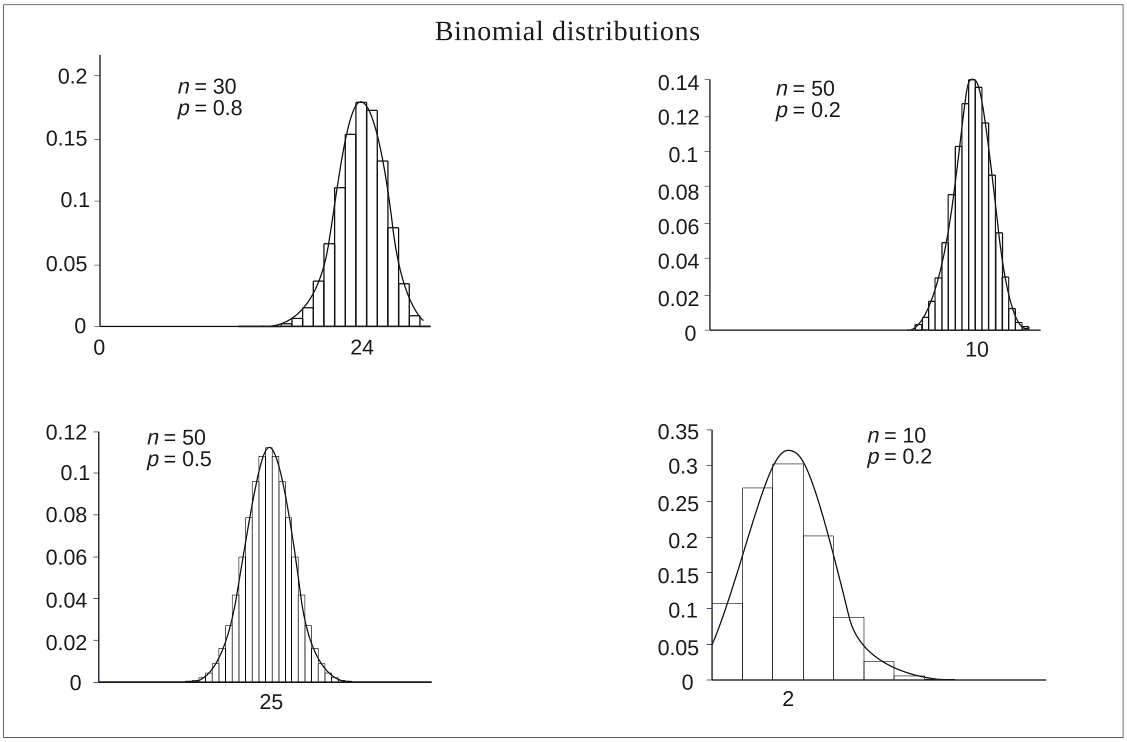
$$= 1 - 0.97381$$

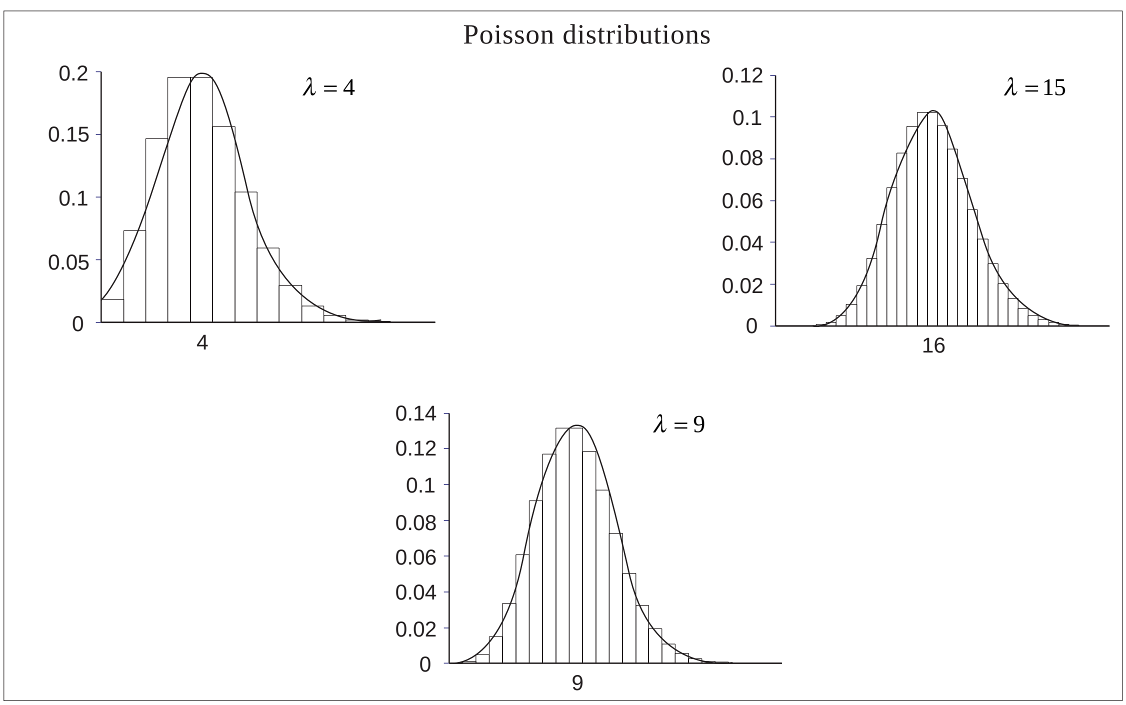
$$= 0.02619$$

(Note that 14.5 was used since **less** than 15 is required.)

Knowing when to use the normal distribution is important. Remember that it is only an approximation and if a simple calculation or tables will give the answer, this should be used.

You may have access to a computer package which can draw histograms of binomial and Poisson distributions for different n, p and λ, and overlay a normal distribution. The following diagrams show this for different cases.





Activity 3

Check that the diagrams illustrate that

(a) for a binomial distribution, if p is close to 0.5, the normal is a good approximation even for quite small n. However, if p is small or large, then a larger value of n will be required for the approximation to be good;

(If $n > 30$, $np > 5$, $nq > 5$, then this is generally regarded as a satisfactory set of circumstances to use a normal approximation.)

(b) for a Poisson distribution, the larger n is the better the approximation.

($\lambda > 20$ is usually regarded as a necessary condition to use a normal approximation.)

To summarise, including the use of the Poisson to approximate to the binomial,

Distribution	Conditions for using approximation	Approximating distribution
$X \sim B(n, p)$	n large (say >50) and p small (say <0.1)	$X \sim Po(n\,p)$
$X \sim B(n, p)$	p close to $\frac{1}{2}$ and $n > 10$ or p moving away from $\frac{1}{2}$ and $n > 30$	$X \sim N(n\,p, n\,p\,q)$ $(q = 1 - p)$
$X \sim Po(\lambda)$	$\lambda > 20$ (say)	$X \sim N(\lambda, \lambda)$

Example

If $X \sim B(20, 0.4)$, find $P(6 \le X \le 10)$.

Also find approximations to this probability by using the

(a) normal distribution

(b) Poisson distribution.

Solution

$$P(X=6) = {}^{20}C_6 (0.6)^{14}(0.4)^6 = 0.1244$$

Similarly $P(X=7) = 0.1659$

$P(X=8) = 0.1797$

$P(X=9) = 0.1597$

$P(X=10) = 0.1171$

Hence $P(6 \le X \le 10) = 0.747$ to 3 decimal places.

(a) Using a normal distribution,

$X \sim N(np, npq)$ where $np = 20 \times 0.4 = 8$

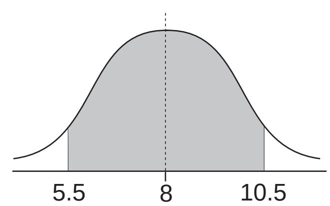
and $npq = 20 \times 0.4 \times 0.6 = 4.8$

So

$X \sim N(8, 4.8)$

and $P(6 \le X \le 10) \rightarrow P(5.5 < X < 10.5)$.

With $Z = \dfrac{X-8}{\sqrt{4.8}}$,



$$\begin{aligned} P(5.5 < X < 10.5) &= \Phi\left(\frac{10.5-8}{\sqrt{4.8}}\right) - \Phi\left(\frac{5.5-8}{\sqrt{4.8}}\right) \\ &= \Phi(1.141) - \Phi(-1.141) \\ &= 2\Phi(1.141) - 1 \\ &\approx 2 \times 0.87286 - 1 \\ &= 0.746 \text{ to 3 decimal places.} \end{aligned}$$

(Note that this is very close to the value found above.)

(b) Using a Poisson distribution

$\lambda = np = 8$

So

$X \sim Po(8)$ and $P(X=x) = e^{-8}\dfrac{8^x}{x!}$

This gives

$$P(X=6) = e^{-8}\frac{8^6}{6!} = 0.1221$$

Similarly $P(X=7) = 0.1396$

$P(X=8) = 0.1396$

$P(X=9)=0.1241$

$P(X=10)=0.0993$

Thus $P(6 \le X \le 10)=0.625$ to 3 decimal places.

(This is a poor approximation, since you should have $n>50$ and $p<\frac{1}{10}$ to use a Poisson approximation.)

Finally you should now be in a position to decide which of the distributions to use in order to model a situation.

Example

Answer the following questions using, in each case, tables of the binomial, Poisson or normal distribution according to which you think is most appropriate.

(a) Cars pass a point on a busy city centre road at an average rate of 7 per five second interval. What is the probability that in a particular five second interval the number of cars passing will be
  (i) 7 or less
  (ii) exactly 7?

(b) Weather records show that for a certain airport during the winter months an average of one day in 25 is foggy enough to prevent landings. What is the probability that in a period of seven winter days landings are prevented on
  (i) 2 or more days?
  (ii) no days?

(c) The working lives of a particular brand of electric light bulb are distributed with mean 1200 hours and standard deviation 200 hours. What is the probability of a bulb lasting more than 1150 hours?

(AEB)

Solution

(a) The Poisson distribution is suitable here since the question concerns a random event that can occur 0, 1, 2, ... times.

The mean value is $x=7$, giving, from tables,

(i) $P(7 \text{ or less})=0.5987$

(ii) $P(7)=P(7 \text{ or less})-P(6 \text{ or less})$

$=0.5987-0.4497$

$=0.149$.

(b) The binomial distribution is a suitable distribution with

$n = 7$ and $p = \frac{1}{25}$. Using tables,

(i) $P(2 \text{ or more}) = 1 - P(1 \text{ or less})$

$$= 1 - 0.9706$$
$$= 0.0294.$$

(ii) $P(\text{no days}) = \left(\frac{24}{25}\right)^7$

$$\approx 0.7514.$$

(c) The normal distribution is the model to use here, although the 'working lives' are not necessarily normal; so assume X, the working life, is distributed

$$X \sim N\left(1200,\ 200^2\right)$$

and

$$P(X > 1150) = 1 - P(X \le 1150)$$

$$= 1 - \Phi\left(\frac{1150 - 1200}{200}\right)$$
$$= 1 - \Phi(-0.25)$$
$$= \Phi(0.25)$$
$$= 0.59871.$$

Exercise 8E

1. The probability of someone smoking is about 0.4. What is the probability that:

   (a) in a group of 50 people more than half of them smoke;

   (b) in a group of 150, less than 50 of them smoke?

2. It is known nationally that support for the Story party is 32% from election results. In a survey carried out on 200 voters what is the probability that more than 80 of them are Story supporters?

3. A manufacturer knows from experience that his machines produce defects at a rate of 5%. In a day's production of 500 items 40 defects are produced. The Production Manager says this is not surprising. Is there evidence to support this?

4. Tickets for a concert are sold according to a Poisson distribution with mean 30 per day. What are the probabilities that:

   (a) less than 20 tickets are sold in one day;

   (b) all 180 tickets are sold in a five day working week?

5. Parts for a washing machine are known to have a weekly demand according to a Poisson distribution mean 20. How many parts should be stocked to ensure that a shop only runs out of parts on 1 in 20 weeks?

8.6 A very important application of the normal

Most modern calculators have a random number generator. The numbers produced generally follow a rectangular distribution in the range 0.000 to 0.999. These should therefore have mean 0.5, variance $0.083\ (\sigma = 0.289)$. (See Section 7.6)

Activity 4

Generate 10 random numbers and put them straight into the statistical function of your calculator. Write down *x*, the mean of your sample.

Repeat this 20 times and write down the means of the samples (remember to clear the statistical memories each time).

Plot these twenty results on normal probability paper and find the mean and SD of the sample means.

You should find that the twenty values are roughly normal, with mean, not suprisingly, 0.5 and SD 0.1. The SD has been decreased by a factor equivalent to the square root of the size of the sample, i.e. $\sqrt{10} = 3.16$.

This is the basis of a very important theorem, called the **Central Limit Theorem**. This says that, irrespective of the original distribution, sample means are normally distributed about the original distribution mean with 'standard error' equal to $\frac{\sigma}{\sqrt{n}}$, σ being the original SD and *n* the sample size. This will be explained in more detail in the next chapter.

8.7 Miscellaneous Exercises

1. The masses of plums from a certain orchard have mean 24g and standard deviation 5g. The plums are graded small, medium or large. All plums over 28g in mass are regarded as large and the rest equally divided between small and medium. Assuming a normal distribution find:

   (a) the proportion of plums graded large;

   (b) the upper limit of the masses of the plums in the small grade. (AEB)

2. A student is doing a project on the hire of videos from a local shop. She finds that the daily demand for videos is approximately normal, with mean 50 and SD 10.

   (a) What is the probability of more than 65 videos being hired on a particular day?

   (b) The shop is considering stopping the hire as it is uneconomical and decides that if demand is less than 40 on more than 3 days out of the next 7 it will do this.

   How likely is this to happen?

   (c) The student reckons that with a wider range of videos, demand would increase by 25% on average with no effect on the SD.

   What is the probability of more than 65 videos being hired if this happens?

3. A Dungeons & Dragons player is suspicious of a new die he has bought. He rolls the die 200 times and says he will throw it away if he gets more than 40 sixes. What is the probability of this happening with a fair die?

   A friend who is a Statistics student suggests that it would be better not to use 40 but to take a figure which a fair die would only exceed 5% of the time. What would this figure be?

4. In a survey of heights it was found that, of males in the 16 - 19 year old age group, 25% were taller than 178.8 cm and 10% were smaller than 165.4. Use this information to find the mean and SD of the distribution assuming it to be normal.

   What is the likelihood of a male in this age group being more than 183 cm (6 feet) tall?

5. Henri de Lade regularly travels from his home in the suburbs to his office in Paris. He always tries to catch the same train, the 08.05, from his local station. He walks to the station from his home in such a way that his arrival times form a normal distribution with mean 08.00 hours and SD 6 minutes.

   (a) Assuming that his train always leaves on time, what is the probability that, on any given day, Henri misses his train?

   (b) If Henri visits his office in this way 5 days each week and if his arrival times at the station each day are independent, what is the probability that he misses his train once, and only once, in a given week?

   (c) Henri visits his office 46 weeks every year. Assuming that there are no absences during this time, what is the probability that he misses his train less than 35 times in the year? (AEB)

6. The weights of pieces of home made fudge are normally distributed with mean 34 g and standard deviation 5 g.

   (a) What is the probability that a piece selected at random weighs more than 40g?

   (b) For some purposes it is necessary to grade the pieces as small, medium or large. It is decided to grade all pieces weighing over 40 g as large and to grade the heavier half of the remainder as medium. The rest will be graded as small. What is the upper limit of the small grade? (AEB)

7. Yuk Ping belongs to an athletics club. In javelin throwing competitions her throws are normally distributed with mean 41.0 m and standard deviation 2.0 m.

   (a) What is the probability of her throwing between 40 m and 46 m?

   (b) What distance will be exceeded by 60% of her throws?

   Gwen belongs to the same club. In competitions 85% of her javelin throws exceed 35 m and 70% exceed 37.5 m. Her throws are normally distributed.

   (c) Find the mean and standard deviation of Gwen's throws, each correct to two significant figures.

   (d) The club has to choose one of these two athletes to enter a major competition. In order to qualify for the final round it is necessary to achieve a throw of at least 48 m in the preliminary rounds. Which athlete should be chosen and why? (AEB)

8. Describe the main features of a normal distribution.

   A company has two machines cutting cylindrical corks for wine bottles. The diameters of corks produced by each machine are normally distributed. The specification requires corks with diameters between 2.91 cm and 3.12 cm.

   Corks cut on Machine A have diameters with a mean 3.03 cm and standard deviation 0.05 cm.

Calculate the percentage of corks cut on this machine that

(a) are rejected as undersize.

(b) meet the specification.

Machine B cuts corks with a mean diameter of 3.01 cm of which 1.7% are rejected as oversize. Calculate the standard deviation of the diameters of corks cut on Machine B.

Which machine, if either, do you consider to be the better? Explain. (AEB)

9. In parts (a) and (b) of this question use the binomial, Poisson or normal distribution according to which you think is the most appropriate. In each case draw attention to any feature of the data which supports or casts doubt on the suitability of the model you have chosen. Indicate, where appropriate, that you are using one distribution as an approximation to another.

(a) A technician looks after a large number of machines on a night shift. She has to make frequent minor adjustments. The necessity for these occurs at random at a constant average rate of 8 per hour. What is the probability that

(i) in a particular hour she will have to make 5 or fewer adjustments;

(ii) in an eight hour shift she will have to make 70 or more adjustments?

(b) A number of neighbouring allotment tenants bought a large quantity of courgette seeds which they shared between them. Overall 15% failed to germinate. What is the probability that a tenant who planted 20 seeds would have

(i) 5 or more failing to germinate;

(ii) at least 17 germinating? (AEB)