Contents 39

The Normal Distribution

Learning outcomes

In a previous Workbook you learned what a continuous random variable was. Here you will examine the most important example of a continuous random variable: the normal distribution. The probabilities of the normal distribution have to be determined numerically. Tables of such probabilities, which refer to a simplified normal distribution called the standard normal distribution, which has mean 0 and variance 1, will be used to determine probabilities of the general normal distribution. Finally you will learn how to deal with combinations of random variables which is an important statistical tool applicable to many engineering situations.

The Normal Distribution

39.1

Introduction

Mass-produced items should conform to a specification. Usually, a mean is aimed for but due to random errors in the production process a tolerance is set on deviations from the mean. For example if we produce piston rings which have a target mean internal diameter of 45 mm then realistically we expect the diameter to deviate only slightly from this value. The deviations from the mean value are often modelled very well by the **normal distribution**. Suppose we decide that diameters in the range 44.95 mm to 45.05 mm are acceptable, then what proportion of the output is satisfactory? In this Section we shall see how to use the normal distribution to answer questions like this.

Prerequisites

Before starting this Section you should ...

- be familiar with the basic properties of probability
- be familiar with continuous random variables

Learning Outcomes

On completion you should be able to ...

- recognise the shape of the frequency curve for the normal distribution and the standard normal distribution
- calculate probabilities using the standard normal distribution
- recognise key areas under the frequency curve

1. The normal distribution

The normal distribution is the most widely used model for the distribution of a random variable. There is a very good reason for this. Practical experiments involve measurements and measurements involve errors. However you go about measuring a quantity, inaccuracies of all sorts can make themselves felt. For example, if you are measuring a length using a device as crude as a ruler, you may find errors arising due to:

- the calibration of the ruler itself;
- parallax errors due to the relative positions of the object being measured, the ruler and your eye;
- rounding errors;
- 'guesstimation' errors if a measurement is between two marked lengths on the ruler.
- mistakes.

If you use a meter with a digital readout, you will avoid some of the above errors but others, often present in the design of the electronics controlling the meter, will be present. Errors are unavoidable and are usually the sum of several factors. The behaviour of variables which are the sum of several other variables is described by a very important and powerful result called the Central Limit Theorem which we will study later in this Workbook. For now we will quote the result so that the importance of the normal distribution will be appreciated.

The central limit theorem

Let X be the sum of n independent random variables $X_i, i = 1, 2, \ldots n$ each having a distribution with mean μ_i and variance $\sigma_i^2 \quad (\sigma_i^2 < \infty)$, respectively, then the distribution of X has expectation and variance given by the expressions

$$\mathsf{E}(X) = \sum_{i=1}^{n} \mu_i \qquad \text{and} \qquad \mathsf{V}(X) = \sum_{i=1}^{n} \sigma_i^2$$

and becomes **normal** as $n \to \infty$.

Essentially we are saying that a quantity which represents the combined effect of a number of variables will be approximately normal no matter what the original distributions are provided that $\sigma^2 < \infty$. This statement is true for the vast majority of distributions you are likely to meet in practice. This is why the normal distribution is crucially important to engineers. A quotation attributed to Prof. G. Lippmann, (1845-1921, winner of the Nobel prize for Physics in 1908) 'Everybody believes on the law of errors, experimenters because they think it is a mathematical theorem andmathematicians because they think it is an experimental fact.'

You may think that anything you measure follows an approximate normal distribution. Unfortunately this is not the case. While the heights of human beings follow a normal distribution, weights do not. Heights are the result of the interaction of many factors (outside one's control) while weights principally depend on lifestyle (including how much and what you eat and drink!) In practice, it is found that weight is skewed to the right but that the square root of human weights is approximately normal.

The probability density function of a normal distribution with mean μ and variance σ^2 is given by the formula

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-(x-\mu)^2/2\sigma^2}$$

This curve is always bell-shaped with the centre of the bell located at the value of μ. See Figure 1. The height of the bell is controlled by the value of σ. As with all normal distribution curves it is symmetrical about the centre and decays as $x \to \pm\infty$. As with any probability density function the area under the curve is equal to 1.

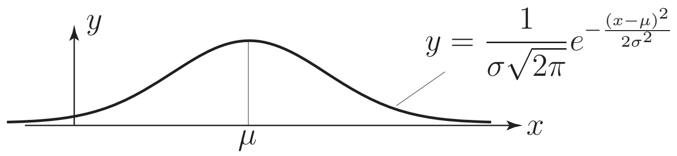


Figure 1

A normal distribution is completely defined by specifying its mean (the value of μ) and its variance (the value of σ^2.) The normal distribution with mean μ and variance σ^2 is written $N(\mu, \sigma^2)$. Hence the distribution $N(20, 25)$ has a mean of 20 and a standard deviation of 5; remember that the second parameter is the variance which is the square of the standard deviation.

Key Point 1

A normal distribution has mean μ and variance σ^2. A random variable X following this distribution is usually denoted by $N(\mu, \sigma^2)$ and we often write

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

Clearly, since μ and σ^2 can both vary, there are infinitely many normal distributions and it is impossible to give tabulated information concerning them all.

For example, if we produce piston rings which have a target mean internal diameter of 45 mm then we may realistically expect the actual diameter to deviate from this value. Such deviations are well-modelled by the normal distribution. Suppose we decide that diameters in the range 44.95 mm to 45.05 mm are acceptable, we may then ask the question 'What proportion of our manufactured output is satisfactory?'

Without tabulated data concerning the appropriate normal distribution we cannot easily answer this question (because the integral used to calculate areas under the normal curve is intractable.)
Since tabulated data allow us to apply the distribution to a wide variety of statistical situations, and we cannot tabulate all normal distributions, we tabulate only one - the standard normal distribution - and convert all problems involving the normal distribution into problems involving the standard normal distribution.

2. The standard normal distribution

At this stage we shall, for simplicity, consider what is known as a standard normal distribution which is obtained by choosing particularly simple values for μ and σ.

Key Point 2

The **standard normal distribution** has a mean of zero and a variance of one.

In Figure 2 we show the graph of the standard normal distribution which has probability density function $y = \dfrac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$

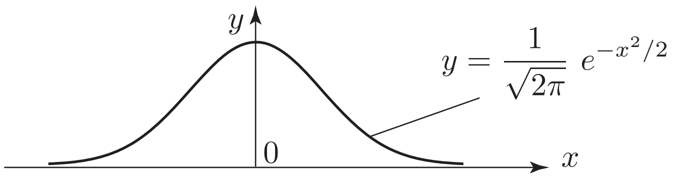


Figure 2: The standard normal distribution curve

The result which makes the standard normal distribution so important is as follows:

Key Point 3

If the behaviour of a continuous random variable X is described by the distribution $N(\mu, \sigma^2)$ then the behaviour of the random variable $Z = \dfrac{X-\mu}{\sigma}$ is described by the standard normal distribution $N(0,1)$.

We call Z the **standardised normal variable** and we write

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

Example 1

If the random variable X is described by the distribution $N(45, 0.000625)$ then what is the transformation required to obtain the standardised normal variable?

Solution

Here, $\mu = 45$ and $\sigma^2 = 0.000625$ so that $\sigma = 0.025$. Hence $Z = (X - 45)/0.025$ is the required transformation.

Example 2

When the random variable $X \sim N(45, 0.000625)$ takes values between 44.95 and 45.05, between which values does the random variable Z lie?

Solution

When $X = 45.05$, $Z = \dfrac{45.05 - 45}{0.025} = 2$

When $X = 44.95$, $Z = \dfrac{44.95 - 45}{0.025} = -2$

Hence Z lies between -2 and 2.

Task

The random variable X follows a normal distribution with mean 1000 and variance 100. When X takes values between 1005 and 1010, between which values does the standardised normal variable Z lie?

Your solution

Answer

The transformation is $Z = \dfrac{X - 1000}{10}$.

When $X = 1005, Z = \dfrac{5}{10} = 0.5$

When $X = 1010,\ Z = \dfrac{10}{10} = 1.$

Hence Z lies between 0.5 and 1.

3. Probabilities and the standard normal distribution

Since the standard normal distribution is used so frequently a table of values has been produced to help us calculate probabilities - located at the end of the Workbook. It is based upon the following diagram:

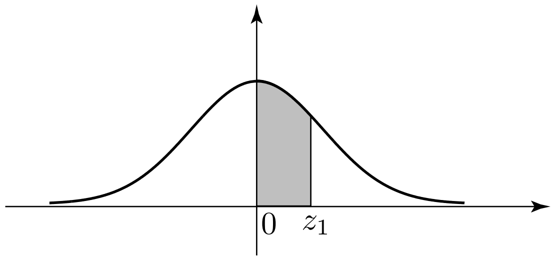


Figure 3

Since the total area under the curve is equal to 1 it follows from the symmetry in the curve that the area under the curve in the region $Z > 0$ is equal to 0.5. In Figure 3 the shaded area is the probability that Z takes values between 0 and z_1. When we 'look-up' a value in the table we obtain the value of the shaded area.

Example 3

What is the probability that Z takes values between 0 and 1.9? (Refer to the table of normal probabilities at the end of the Workbook.)

Solution

The row beginning '1.9' and the column headed '0' is the appropriate choice and its entry is 4713. This is to be read as 0.4713 (we omitted the '0.' in each entry for clarity.) The interpretation is that the probability that Z takes values between 0 and 1.9 is 0.4713.

Example 4

What is the probability that Z takes values between 0 and 1.96?

Solution

This time we want the row beginning 1.9 and the column headed '6'.

The entry is 4750 so that the required probability is 0.4750.

Example 5

What is the probability that Z takes values between 0 and 1.965?

Solution

There is no entry corresponding to 1.965 so we take the average of the values for 1.96 and 1.97. (This linear interpolation is not strictly correct but is acceptable.)

The two values are 4750 and 4756 with an average of 4753. Hence the required probability is 0.4753.

Task

What are the probabilities that Z takes values between

(a) 0 and 2 (b) 0 and 2.3 (c) 0 and 2.33 (d) 0 and 2.333?

Your solution

Answer

(a) The entry is 4772; the probability is 0.4772.

(b) The entry is 4893; the probability is 0.4893.

(c) The entry is 4901; the probability is 0.4901.

(d) The entry for 2.33 is 4901, that for 2.34 is 4904.

Linear interpolation gives a value of $4901 + 0.3(4904 - 4901)$ i.e. about 4902; the probability is 0.4902.

Note from Table 1 that as Z increases from 0 the entries increase, rapidly at first and then more slowly, toward 5000 i.e. a probability of 0.5. This is consistent with the shape of the curve.

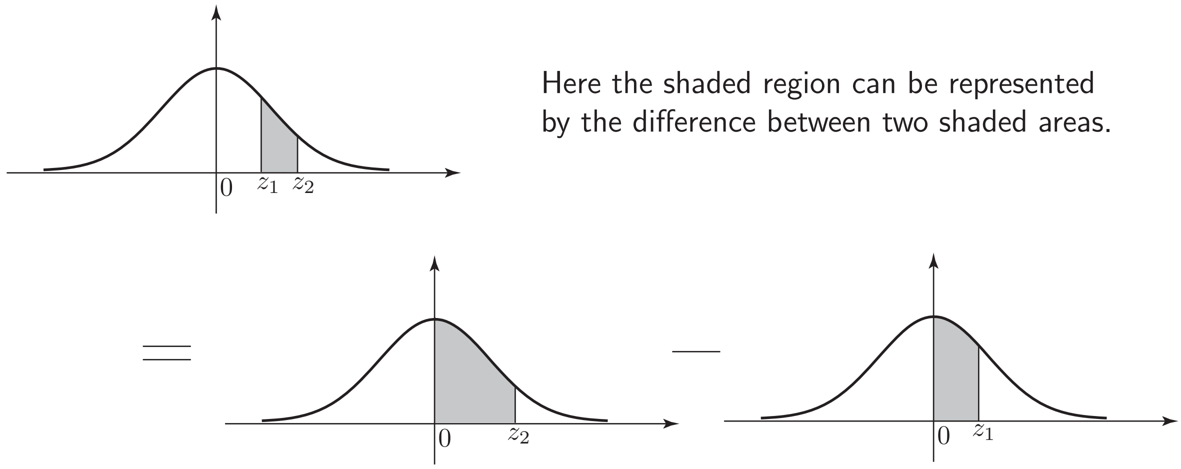
After $Z = 3$ the increase is quite slow so that we tabulate entries for values of Z rising by increments of 0.1 instead of 0.01 as in the rest of Table 1.

4. Calculating other probabilities

In this Section we see how to calculate probabilities represented by areas other than those of the type shown in Figure 3.

Case 1

Figure 4 illustrates what we do if both Z values are positive. By using the properties of the standard normal distribution we can organise matters so that any required area is always of 'standard form'.



Here the shaded region can be represented by the difference between two shaded areas.

Figure 4

Example 6

Find the probability that Z takes values between 1 and 2.

Solution

Using Table 1:

P$(Z = z_2)$ i.e. P$(Z = 2)$ is 0.4772

P$(Z = z_1)$ i.e. P$(Z = 1)$ is 0.3413.

Hence P$(1 < Z < 2) = 0.4772 - 0.3413 = 0.1359$

Remember that with a continuous distribution, P$(Z = 1)$ is meaningless (will have zero probability) so that P$(1 \leq Z \leq 2)$ is interpreted as P$(1 < Z < 2)$.

Case 2

The following diagram illustrates the procedure to be followed when finding probabilities of the form $\mathsf{P}(Z > z_1)$.

This time the shaded area is the difference between the right-hand half of the total area and an area which can be read off from Table 1.

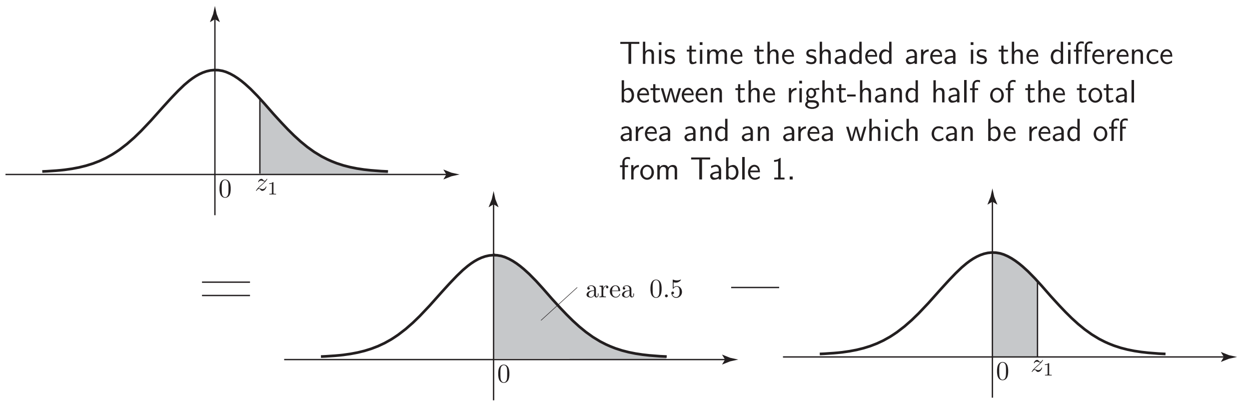


Figure 5

Example 7

What is the probability that $Z > 2$?

Solution

$\mathsf{P}(0 < Z < 2) = 0.4772$ (from Table 1). Hence the probability is $0.5 - 0.4772 = 0.0228$.

Case 3

Here we consider the procedure to be followed when calculating probabilities of the form $\mathsf{P}(Z < z_1)$. Here the shaded area is the sum of the left-hand half of the total area and a 'standard' area.

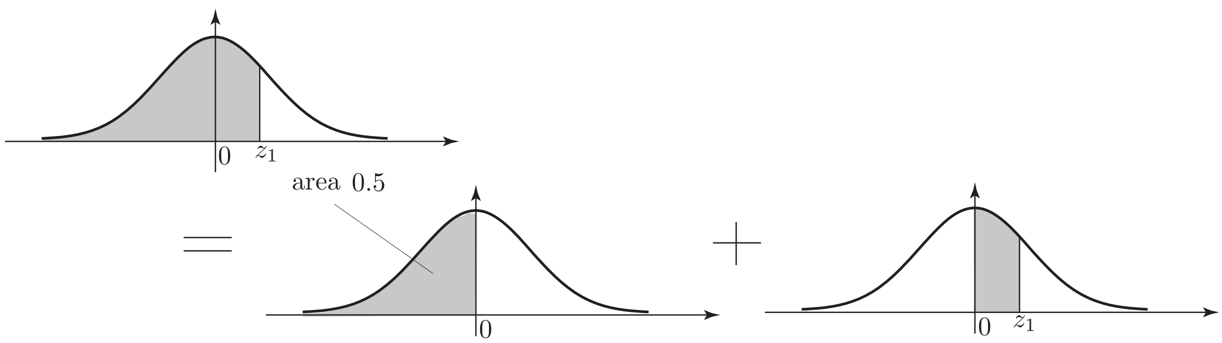


Figure 6

Example 8

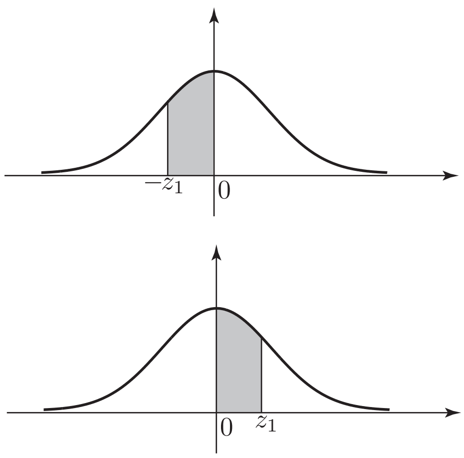
What is the probability that $Z < 2$?

Solution

$\mathsf{P}(Z < 2) = 0.5 + 0.4772 = 0.9772.$

Case 4

Here we consider what needs to be done when calculating probabilities of the form $\mathsf{P}(-z_1 < Z < 0)$ where z_1 is positive. This time we make use of the symmetry in the standard normal distribution curve.



By symmetry this shaded area is equal in value to the one above.

Figure 7

Example 9

What is the probability that $-2 < Z < 0$?

Solution

The area is equal to that corresponding to $\mathsf{P}(0 < Z < 2) = 0.4772.$

Case 5

Finally we consider probabilities of the form P$(-z_2 < Z < z_1)$. Here we use the sum property and the symmetry property.

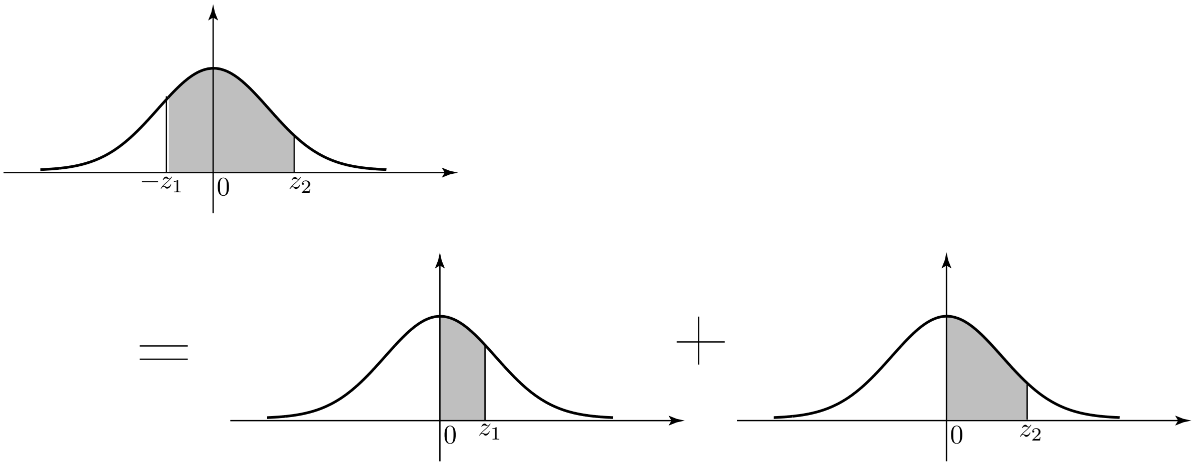


Figure 8

Example 10

What is the probability that $-1 < Z < 2$?

Solution

$$\begin{aligned} \text{P}(-1 < Z < 0) = \text{P}(0 < Z < 1) &= 0.3413 \\ \text{P}(0 < Z < 2) &= 0.4772 \end{aligned}$$

Hence the required probability P$(-1 < Z < 2)$ is 0.8185.

Other cases can be handed by a combination of the ideas already used.

Task

Find the following probabilities.

(a)	$\mathsf{P}(0 < Z < 1.5)$	(b)	$\mathsf{P}(Z > 1.8)$
(c)	$\mathsf{P}(1.5 < Z < 1.8)$	(d)	$\mathsf{P}(Z < 1.8)$
(e)	$\mathsf{P}(-1.5 < Z < 0)$	(f)	$\mathsf{P}(Z < -1.5)$
(g)	$\mathsf{P}(-1.8 < Z < -1.5)$	(h)	$\mathsf{P}(-1.5 < Z < 1.8)$

(A simple sketch of the standard normal curve will help.)

Your solution

Answer

(a) 0.4332 (direct from Table 1)

(b) $0.5 - 0.4641 = 0.0359$

(c) $\mathsf{P}(0 < Z < 1.8) - \mathsf{P}(0 < Z < 1.5) = 0.4641 - 0.4332 = 0.0309$

(d) $0.5 + 0.4641 = 0.9641$

(e) $\mathsf{P}(-1.5 < Z < 0) = \mathsf{P}(0 < Z < 1.5) = 0.4332$

(f) $\mathsf{P}(Z < -1.5) = \mathsf{P}(Z > 1.5) = 0.5 - 0.4332 = 0.0668$

(g) $\mathsf{P}(-1.8 < Z < -1.5) = \mathsf{P}(1.5 < Z < 1.8) = 0.0309$

(h) $\mathsf{P}(0 < Z < 1.5) + \mathsf{P}(0 < Z < 1.8) = 0.8973$

5. The cumulative distribution function

We know that the normal probability density function $f(x)$ is given by the formula

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-(x-\mu)^2/2\sigma^2}$$

and so the cumulative distribution function $F(x)$ is given by the formula

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-(u-\mu)^2/2\sigma^2}du$$

In the case of the cumulative distribution for the standard normal curve, we use the special notation $\Phi(z)$ and, substituting 0 and 1 for μ and σ^2, we obtain

$$\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\mathrm{e}^{-u^2/2}du$$

The shape of the curve is essentially 'S' -shaped as shown in Figure 9. Note that the curve runs from $-\infty$ to $+\infty$. As you can see, the curve approaches the value 1 asymptotically.

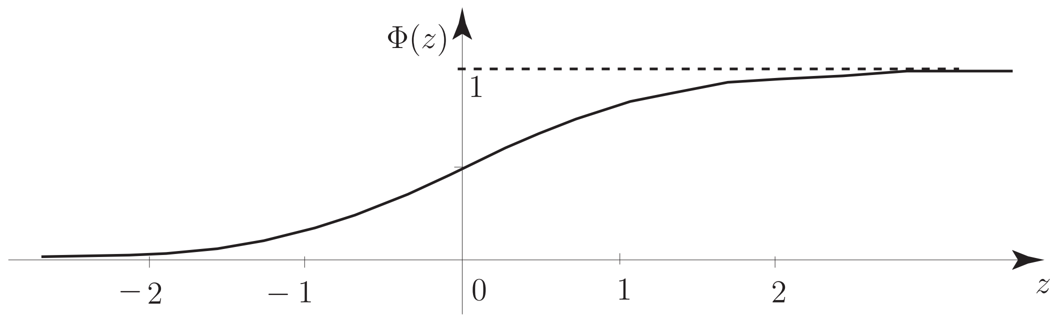


Figure 9

Comparing the integrals

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-(u-\mu)^2/2\sigma^2}du \qquad \text{and} \qquad \Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\mathrm{e}^{-v^2/2}dv$$

shows that

$$v = \frac{u-\mu}{\sigma} \qquad \text{and so} \qquad dv = \frac{du}{\sigma}$$

and $F(x)$ may be written as

$$\begin{aligned} F(x) &= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{(x-\mu)/\sigma}\mathrm{e}^{-v^2/2}\sigma dv \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{(x-\mu)/\sigma}\mathrm{e}^{-v^2/2}dv = \Phi(\frac{x-\mu}{\sigma}) \end{aligned}$$

We already know, from the basic definition of a cumulative distribution function, that

$$\mathrm{P}(a < X < b) = F(b) - F(a)$$

so that we may write the probability statement above in terms of $\Phi(z)$ as

$$\mathrm{P}(a < X < b) = F(b) - F(a) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma}).$$

The value of $\Phi(z)$ is measured from $z = -\infty$ to any ordinate $z = z_1$ and represents the probability $\mathsf{P}(Z < z_1)$.
The values of $\Phi(z)$ start as shown below:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0.1	.5398	5438	5478	5517	5577	5596	5636	5675	5714	5753
0.2	.5793	5832	5871	5909	5948	5987	6026	6064	6103	6141

You should compare the values given here with the values given for the normal probability integral (Table 1 at the end of the Workbook). Simply adding 0.5 to the values in the latter table gives the values of $\Phi(z)$. You should also note that the diagrams shown at the top of each set of tabulated values tells you whether you are looking at the values of $\Phi(z)$ or the values of the normal probability integral.

Exercises

1. If a random variable X has a standard normal distribution find the probability that it assumes a value:

   (a) less than 2.00

   (b) greater than 2.58

   (c) between 0 and 1.00

   (d) between -1.65 and -0.84

2. If X has a standard normal distribution find k in each of the following cases:

   (a) $\mathsf{P}(X < k) = 0.4$

   (b) $\mathsf{P}(X < k) = 0.95$

   (c) $\mathsf{P}(0 < X < k) = 0.1$

Answers

1 (a) 0.9772 (b) 0.0049 (c) 0.3413 (d) 0.1510

2 (a) -0.2533 (b) 1.6450 (c) 0.2533

6. Applications of the normal distribution

We have, in the previous subsection, noted that the probability density function of a normal distribution X is

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

This curve is always 'bell-shaped' with the centre of the bell located at the value of μ. The height of the bell is controlled by the value of σ. See Figure 10.

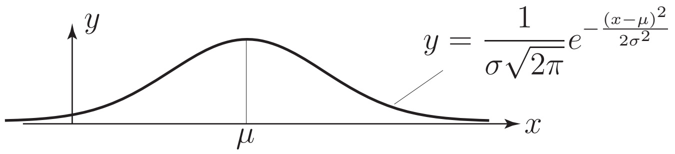


Figure 10

We now show, by example, how probabilities relating to a general normal distribution X are determined. We will see that being able to calculate the probabilities of a standard normal distribution Z is crucial in this respect.

Example 11

Given that the variate X follows the normal distribution $X \sim N(151, 15^2)$, calculate:

(a) P$(120 \leq X \leq 155)$; (b) P$(X \geq 185)$

Solution

The transformation used in this problem is $Z = \dfrac{X - \mu}{\sigma} = \dfrac{X - 151}{15}$

(a)

$$\begin{aligned} \text{P}(120 \leq X \leq 155) &= \text{P}\left(\frac{120-151}{15} \leq Z \leq \frac{155-151}{15}\right) \\ &= \text{P}(-2.07 \leq Z \leq 0.27) \\ &= 0.4808 + 0.1064 = 0.5872 \end{aligned}$$

(b)

$$\begin{aligned} \text{P}(X \geq 185) &= \text{P}\left(Z \geq \frac{185-151}{15}\right) \\ &= \text{P}(Z \geq 2.27) \\ &= 0.5 - 0.4884 = 0.0116 \end{aligned}$$

We note that, as for any continuous random variable, we can only calculate the probability that

- X lies between two given values;
- X is greater than a given value;
- X is less that a given value.

rather than for individual values.

Task

A worn, poorly set-up machine is observed to produce components whose length X follows a normal distribution with mean 20 cm and variance 2.56 cm Calculate:

(a) the probability that a component is at least 24 cm long;
(b) the probability that the length of a component lies between 19 and 21 cm.

Your solution

Answer

The transformation used is $Z = \dfrac{X-20}{1.6}$ giving

$$\mathrm{P}(X \geq 24) = \mathrm{P}(Z \geq \frac{24-20}{1.6}) = \mathrm{P}(Z \geq 2.5) = 0.5 - 0.4938 = 0.0062$$

and

$$\mathrm{P}(19 < X < 21) = \mathrm{P}(\frac{19-20}{1.6} < Z < \frac{21-20}{1.6}) = \mathrm{P}(-0.625 < Z < 0.625) = 0.4681$$

Example 12

Piston rings are mass-produced. The target internal diameter is 45 mm but records show that the diameters are normally distributed with mean 45 mm and standard deviation 0.05 mm. An acceptable diameter is one within the range 44.95 mm to 45.05 mm. What proportion of the output is unacceptable?

Solution

There are many words in the statement of the problem; we must read them carefully to extract the necessary information. If X is the diameter of a piston ring then $X \sim N(45, (0.05)^2)$.

The transformation is $Z = \dfrac{X - \mu}{\sigma} = \dfrac{X - 45}{0.05}$.

The upper limit of acceptability is $x_2 = 45.05$ so that $z_2 = (45.05 - 45)/0.05 = 1$.

The lower limit of acceptability is $x_1 = 44.95$ so that $z_1 = (44.95 - 45)/0.05 = -1$.

The range of 'acceptable' Z values is therefore -1 to 1. Figure 11 below.

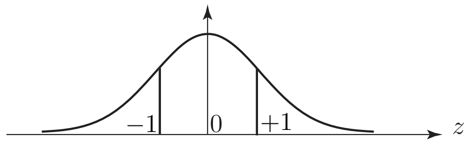


Figure 11

Using the symmetry of the curves P$(-1 < Z < 1) = 2 \times$ P$(0 < Z < 1) = 2 \times 0.3413 = 0.6826$.

Thus the proportion of unacceptable items is $1 - 0.6826 = 0.3174$, or 31.74%.

Example 13

If the standard deviation is halved by improved production practices what is now the proportion of unacceptable items?

Solution

Now $\sigma = 0.025$ so that:

$$z_2 = \frac{45.05 - 45}{0.025} = 2 \qquad \text{and} \qquad z_1 = -2$$

Then P$(-2 < Z < 2) = 2 \times$ P$(0 < Z < 2) = 2 \times 0.4772 = 0.9544$. Hence the proportion of unacceptable items is reduced to $1 - 0.9544 = 0.0456$ or 4.56%.
We observe that less of the area under the curve now lies outside the interval $(44.95, 45.05)$.

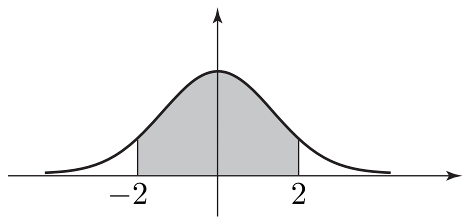


Figure 12

Task

The resistance of a strain gauge is normally distributed with a mean of 100 ohms and a standard deviation of 0.2 ohms. To meet the specification, the resistance must be within the range 100 ± 0.5 ohms.

(a) What percentage of gauges are unacceptable?

First, state the upper and lower limits of acceptable resistance and find the $Z-$values which correspond:

Your solution

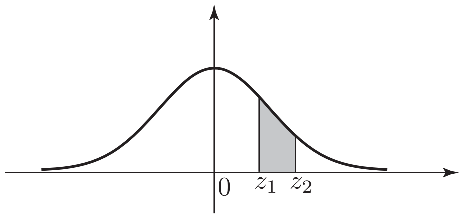
Answer

$x_1 = 99.5, \quad x_2 = 100.5 \qquad Z = \dfrac{X - 100}{0.2} \qquad \left\{(0.2)^2 = 0.04\right\}$ so that $z_1 = -2.5$ and $z_2 = 2.5$

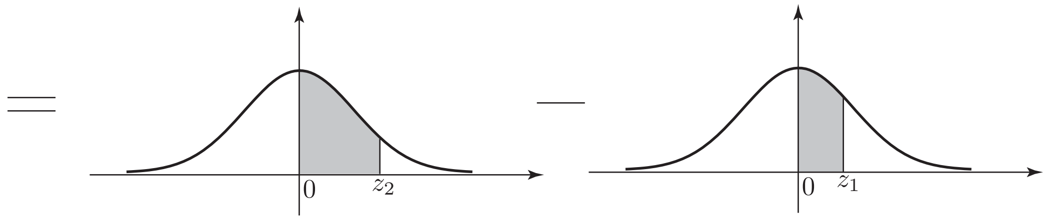
Now, using a suitable sketch, calculate the probability that $z_1 < Z < z_2$:

Your solution

Answer



Here the shaded region can be represented by the difference between two shaded areas.



The shaded area (see diagram) is 0.4938 (from the table of values on page 15). Using symmetry,

$$\begin{aligned} \text{P}(-2.5 < Z < 2.5) &= 2 \times 0.4938 \\ &= 0.9876. \end{aligned}$$

Hence the proportion of acceptable gauges is 98.76%.

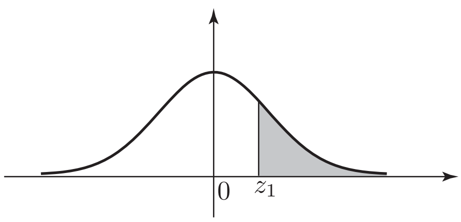
Therefore the proportion of unacceptable gauges is 1.24%.

(b) To what value must the standard deviation be reduced if the proportion of unacceptable gauges is to be no more than 0.2$\%$?

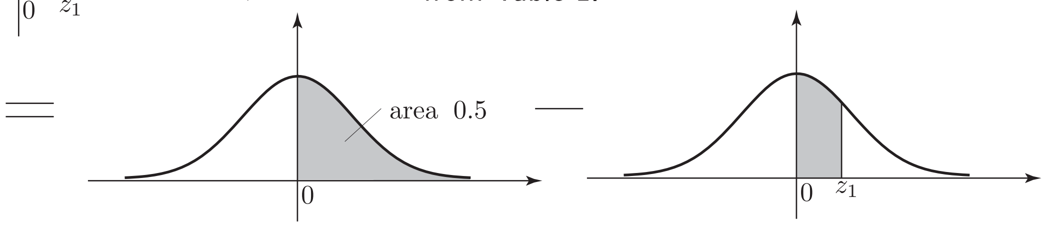
First sketch the standard normal curve marking on it the lower and upper values z_1 and z_2 and appropriate areas:

Your solution

Answer



This time the shaded area is the difference between the right-hand half of the total area and an area which can be read off from Table 1.



Now use the Table to find z_2, and hence write down the value of z_1:

Your solution

Answer

$z_2 = 3.1$ so that $z_1 = -3.1$

Finally, rewrite $Z = \dfrac{X - \mu}{\sigma}$ to make σ the subject. Put in values for $z_2,\ x_2$ and μ hence evaluate σ:

Your solution

Answer

$\sigma = \dfrac{X - \mu}{Z} = \dfrac{100.5 - 100}{3.1} = 0.16$ (2 d.p.)

7. Probability intervals - standard normal distribution

We use probability models to make predictions in situations where there is not sufficient data available to make a definite statement. Any statement based on these models carries with it a **risk** of being proved incorrect by events. Notice that the normal probability curve extends to infinity in both directions. **Theoretically** any value of the normal random variable is possible, although, of course, values far from the mean position (zero) are very unlikely.

Consider the diagram in Figure 13:

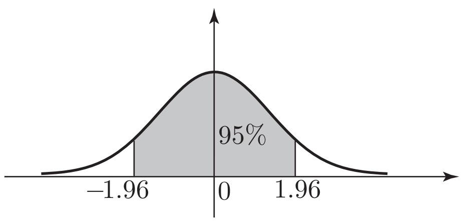


Figure 13

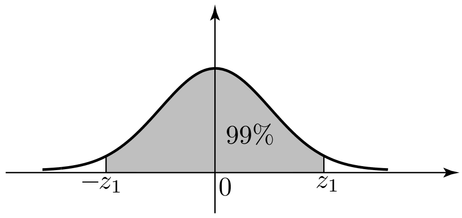
The shaded area is 95% of the total area. If we look at the entry in Table 1 (at the end of the Workbook) corresponding to $Z = 1.96$ we see the value 4750. This means that the probability of Z taking a value between 0 and 1.96 is 0.475. By symmetry, the probability that Z takes a value between -1.96 and 0 is also 0.475. Combining these results we see that

$$\text{P}(-1.96 < Z < 1.96) = 0.95 \text{ or } 95\%$$

We say that the 95% probability interval for Z (about its mean of 0) is $(-1.96, 1.96)$. It follows that there is a 5% chance that Z lies outside this interval.

Task

Find the 99% probability interval for Z about its mean, i.e. the value of z_1 in the diagram:



The shaded area is 99% of the total area

First, note that 99% corresponds to a probability of 0.99. Find z_1 such that

$$\mathrm{P}(0 < Z < z_1) = \frac{1}{2} \times 0.99 = 0.495 :$$

Your solution

Answer
We look for a table value of 4950. The nearest we get is 4949 and 4951 corresponding to $Z = 2.57$ and $Z = 2.58$ respectively. We choose $Z = 2.58$.

Now quote the 99% probability interval:

Your solution

Answer
$(-2.58,\ 2.58)$ or $-2.58 < Z < 2.58$.

Notice that the risk of Z lying outside this wider interval is reduced to 1%.

Task

Find the value of Z

(a) which is exceeded on 5% of occasions

(b) which is exceeded on 99% of occasions.

Your solution

Answer

(a) The value is z_1, where $\mathrm{P}(Z > z_1) = 0.05$. Hence $\mathrm{P}(0 < Z < z_1) = 0.5 - 0.05 = 0.45$ This corresponds to a table entry of 4500. The nearest values are 4495 ($Z = 1.64$) and 4505 ($Z = 1.65$).

Hence the required value is $Z_1 = 1.65$.

(b) Values less than z_1 occur on 1% of occasions. By symmetry values greater than $(-z_1)$ occur on 1% of occasions so that $\mathrm{P}(0 < z < -z_1) = 0.49$. The nearest table corresponding to 4900 is 4901 ($Z = 2.33$).

Hence the required value is $z_1 = -2.33$.

8. Probability intervals - general normal distribution

We saw in subsection 3 that 95% of the area under the standard normal curve lay between $z_1 = -1.96$ and $z_2 = 1.96$. Using the formula $Z = \dfrac{X - \mu}{\sigma}$ in the re-arrangement $X = \mu + Z\sigma$. We can see that 95% of the area under the general normal curve lies between $x_1 = \mu - 1.96\sigma$ and $x_2 = \mu + 1.96\sigma$.

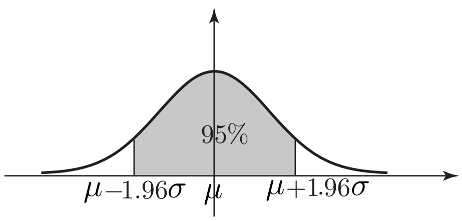


Figure 14

Example 14

Suppose that the internal diameters of mass-produced pipes are normally distributed with mean 50 mm and standard deviation 2 mm. What are the 95% probability limits on the internal diameter of a single pipe?

Solution

Here $\mu = 50$ $\sigma = 2$ so that the 95% probability limits are

$$50 \pm 1.96 \times 2 = 50 \pm 3.92\text{mm}$$

i.e. 46.08 mm and 53.92 mm.
The probability interval is (46.08, 53.92).

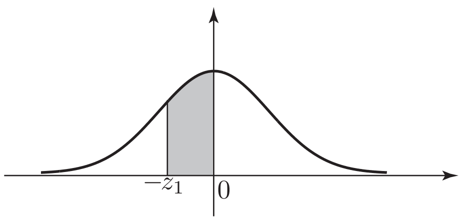
Task

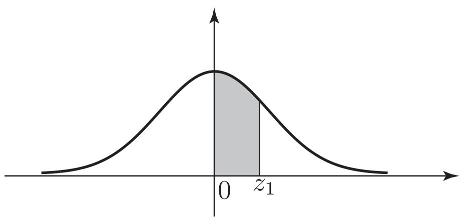
What is the 99% probability interval for the lifetime of a bulb when the lifetimes of such bulbs are normally distributed with a mean of 2000 hours and standard deviation of 40 hours?

First sketch the standard normal curve marking the values z_1, z_2 between which 99% of the area under the curve is located:

Your solution

Answer





By symmetry this shaded area is equal in value to the one above.

Now deduce the corresponding values x_1, x_2 for the general normal distribution:

Your solution

Answer

$x_1 = \mu - 2.58\sigma, \quad x_2 = \mu + 2.58\sigma$

Next, find the values for x_1 and x_2 for the given problem:

Your solution

Answer

$$x_1 = 2000 - 2.58 \times 40 = 1896.8 \text{ hours}$$

$$x_2 = 2000 + 2.58 \times 40 = 2103.2 \text{ hours}$$

Finally, write down the 99% probability interval for the lifetimes:

Your solution

Answer

(1896.8 hours, 2103.2 hours).

The Normal Approximation to the Binomial Distribution 39.2

Introduction

We have already seen that the Poisson distribution can be used to approximate the binomial distribution for large values of n and small values of p provided that the correct conditions exist. The approximation is only of practical use if just a few terms of the Poisson distribution need be calculated. In cases where many - sometimes several hundred - terms need to be calculated the arithmetic involved becomes very tedious indeed and we turn to the normal distribution for help. It is possible, of course, to use high-speed computers to do the arithmetic but the normal approximation to the binomial distribution negates the necessity of this in a fairly elegant way. In the problem situations which follow this introduction the normal distribution is used to avoid very tedious arithmetic while at the same time giving a very good approximate solution.

Prerequisites

Before starting this Section you should ...

- be familiar with the normal distribution and the standard normal distribution
- be able to calculate probabilities using the standard normal distribution

Learning Outcomes

On completion you should be able to ...

- recognise when it is appropriate to use the normal approximation to the binomial distribution
- solve problems using the normal approximation to the binomial distribution.
- interpret the answer obtained using the normal approximation in terms of the original problem

1. The normal approximation to the binomial distribution

A typical problem

An engineering professional body estimates that 75% of the students taking undergraduate engineering courses are in favour of studying of statistics as part of their studies. If this estimate is correct, what is the probability that more than 780 undergraduate engineers out of a random sample of 1000 will be in favour of studying statistics?

Discussion

The problem involves a binomial distribution with a large value of n and so very tedious arithmetic may be expected. This can be avoided by using the normal distribution to approximate the binomial distribution underpinning the problem.

If X represents the number of engineering students in favour of studying statistics, then

$$X \sim B(1000, 0.75)$$

Essentially we are asked to find the probability that X is greater than 780, that is $\mathsf{P}(X > 780)$.
The calculation is represented by the following statement

$$\mathsf{P}(X > 780) = \mathsf{P}(X = 781) + \mathsf{P}(X = 782) + \mathsf{P}(X = 783) + \cdots + \mathsf{P}(X = 1000)$$

In order to complete this calculation we have to find all 220 terms on the right-hand side of the expression. To get some idea of just how big a task this is when the binomial distribution is used, imagine applying the formula

$$\mathsf{P}(X = r) = \frac{n(n-1)(n-2)\ldots(n-r+1)p^r(1-p)^{n-r}}{r(r-1)(r-2)\ldots 3.2.1}$$

220 times! You would have to take $n = 1000,\ p = 0.75$ and vary r from 781 to 1000. Clearly, the task is enormous.

Fortunately, we can approximate the answer very closely by using the normal distribution with the *same mean and standard deviation* as $X \sim B(1000, 0.75)$. Applying the usual formulae for μ and σ we obtain the values $\mu = 750$ and $\sigma = 13.7$ from the binomial distribution.

We now have two distributions, $X \sim B(1000, 0.75)$ and (say) $Y \sim N(750, 13.7^2)$. Remember that the second parameter represents the variance. By doing the appropriate calculations, (this is extremely tedious even for one term!) it can be shown that

$$\mathsf{P}(X = 781) \approx \mathsf{P}(780.5 \leq Y \leq 781.5)$$

This statement means that the probability that $X = 781$ calculated from the binomial distribution $X \sim B(1000, 0.75)$ can be very closely approximated by the area under the normal curve $Y \sim N(750, 13.7^2)$ between 780.5 and 781.5. This relationship is then applied to all 220 terms involved in the calculation.

The result is summarised below:

$$\begin{aligned}
\mathsf{P}(X = 781) &\approx \mathsf{P}(780.5 \leq Y \leq 781.5)\\
\mathsf{P}(X = 782) &\approx \mathsf{P}(781.5 \leq Y \leq 782.5)\\
&\vdots\\
\mathsf{P}(X = 999) &\approx \mathsf{P}(998.5 \leq Y \leq 999.5)\\
\mathsf{P}(X = 1000) &\approx \mathsf{P}(999.5 \leq Y \leq 1000.5)
\end{aligned}$$

By adding these probabilities together we get

$$\begin{aligned}
\mathsf{P}(X > 780) &= \mathsf{P}(X = 781) + \mathsf{P}(X = 782) + \cdots + \mathsf{P}(X = 1000)\\
&\approx \mathsf{P}(780.5 \leq Y \leq 1000.5)
\end{aligned}$$

To complete the calculation we need only to find the area under the curve $Y \sim N(750, 13.7^2)$ between the values 780.5 and 1000.5. This is far easier than completing the 220 calculations suggested by the use of the binomial distribution.

Finding the area under the curve $Y \sim N(750, 13.7^2)$ between the values 780.5 and 1000.5 is easily done by following the procedure used previously. The calculation, using the tables on page 15 and working to three decimal places, is

$$\begin{aligned}
\mathsf{P}(X > 780) &\approx \mathsf{P}(\frac{780.5 - 750}{13.7} \leq Z \leq \frac{1000.5 - 750}{13.7})\\
&= \mathsf{P}(2.23 \leq Z \leq 18.28)\\
&= \mathsf{P}(Z \geq 2.23)\\
&= 0.013
\end{aligned}$$

Notes:

1. Since values as high as 18.28 effectively tell us to find the area to the right of 2.33 (the area to the right of 18.28 is so close to zero as to make no difference) we have

   $$\mathsf{P}(Z \geq 2.23) = 0.0129 \approx 0.013$$

2. The solution given *assumes* that the original binomial distribution can be approximated by a normal distribution. This is not always the case and you must always check that the following conditions are satisfied before you apply a normal approximation. The conditions are:

   - $np > 5$
   - $n(1 - p) > 5$

   You can see that these conditions are satisfied here.

Task

A particular production process used to manufacture ferrite magnets used to operate reed switches in electronic meters is known to give 10% defective magnets on average. If 200 magnets are randomly selected, what is the probability that the number of defective magnets is between 24 and 30?

Your solution

Answer

If X is the number of defective magnets then $X \sim B(200, 0.1)$ and we require

$$\text{P}(24 < X < 30) = \text{P}(25 \leq X \leq 29)$$

Now,

$$\mu = np = 200 \times 0.1 = 20 \quad \text{and} \quad \sigma = \sqrt{np(1-p)} = \sqrt{200 \times 0.1 \times 0.9} = 4.24$$

Note that $np > 5$ and $n(1-p) > 5$ so that approximating $X \sim B(200, 0.1)$ by $Y \sim N(20, 4.24^2)$ is acceptable. We can approximate $X \sim B(200, 0.1)$ by the normal distribution $Y \sim N(20, 4.24^2)$ and use the transformation

$$Z = \frac{Y - 20}{4.24} \sim N(0, 1)$$

so that

$$\begin{aligned}
\text{P}(25 \leq X \leq 29) &\approx \text{P}(24.5 \leq Y \leq 29.5) \\
&= \text{P}\left(\frac{24.5 - 20}{4.24} \leq Z \leq \frac{29.5 - 20}{4.24}\right) \\
&= \text{P}(1.06 \leq Z \leq 2.24) \\
&= 0.4875 - 0.3554 = 0.1321
\end{aligned}$$

Example 15

Overbooking of passengers on intercontinental flights is a common practice among airlines. Aircraft which are capable of carrying 300 passengers are booked to carry 320 passengers. If on average 10% of passengers who have a booking fail to turn up for their flights, what is the probability that at least one passenger who has a booking will end up without a seat on a particular flight?

Solution

Let $p = \text{P(a passenger with a booking, fails to turn up)} = 0.10.$

Then: $q = \text{P(a passenger with a booking, turns up)} = 1 - p = 1 - 0.10 = 0.9$

Let X = number of passengers with a booking who turn up.

As there are 320 bookings, we are dealing with the terms of the binomial expansion of

$$(q+p)^{320} = q^{320} + 320q^{319}p + \frac{320 \times 319}{2!}q^{318}p^2 + \cdots + p^{320}$$

Using this approach is too long to calculate by finding the values term by term. It is easier to switch to the corresponding normal distribution, i.e. that which has the same mean and variance as the binomial distribution above.

$$\text{Mean} = \mu = 320 \times 0.9 = 288$$

$$\text{Variance} = 320 \times 0.9 \times 0.1 = 28.8 \quad \text{so} \quad \sigma = \sqrt{28.8} = 5.37$$

Hence, the corresponding normal distribution is given by $Y \sim N(288, 28.8)$

So that, $\text{P}(X > 300) \approx \text{P}(Y \geq 300.5) \; = \text{P}(Z \geq \dfrac{300.5 - 288}{5.37}) = \text{P}(Z \geq 2.33)$

From Z-tables $\qquad \text{P}(Z \geq 2.33) = 0.0099.$

NB. Continuity correction is needed when changing from the binomial, a discrete distribution, to the normal, a continuous distribution.

Exercises

1. The diameter of an electric cable is normally distributed with mean 0.8 cm and variance 0.0004 cm^2.
   (a) What is the probability that the diameter will exceed 0.81 cm?
   (b) The cable is considered defective if the diameter differs from the mean by more than 0.025 cm. What is the probability of obtaining a defective cable?

2. A machine packs sugar in what are nominally 2 cm kg bags. However there is a variation in the actual weight which is described by the normal distribution.
   (a) Previous records indicate that the standard deviation of the distribution is 0.02 cm kg and the probability that the bag is underweight is 0.01. Find the mean value of the distribution.
   (b) It is hoped that an improvement to the machine will reduce the standard deviation while allowing it to operate with the same mean value. What value standard deviation is needed to ensure that the probability that a bag is underweight is 0.001?

3. Rods are made to a nominal length of 4 cm but in fact the length is a normally distributed random variable with mean 4.01 cm and standard deviation 0.03. Each rod costs 6p to make and may be used immediately if its length lies between 3.98 cm and 4.02 cm. If its length is less than 3.98 cm the rod cannot be used but has a scrap value of 1p. If the length exceeds 4.02 cm it can be shortened and used at a further cost of 2p. Find the average cost per usable rod.

4. A supermarket chain sells its 'own-brand' label instant coffee in packets containing 200 gm of coffee granules. The packets are filled by a machine which is set to dispense fills of 200 gm If fills are normally distributed, about a mean of 200 gm and with a standard deviation of 7 gm, find the number of packets out of a consignment of 1,000 packets that:
   (a) contain more than 215 gm
   (b) contain less than 195 gm
   (c) contain between 190 to 210 gm

   The supermarket chain decides to withdraw all packets with less than a certain weight of coffee. As a result, 40 packets which were in the consignment of 1,000 packets are withdrawn. What is the weight at which the 'line has been drawn' ?

5. The time taken by a team to complete the assembly of an electrical component is found to be normally distributed, about a mean of 110 minutes, and with a standard deviation of 10 minutes.
   (a) Out of a group of 20 teams, how many will complete the assembly:

   (i) within 95 minutes. (ii) in more than 2 hours.

   (b) If the management decides to set a 'cut off' time such that 95% of the teams will have completed the assembly on time, what time limit should be set?

Answers

1. $X \sim N(0.8,\ 0.0004)$

(a) $\mathsf{P}(X > 0.81) = \mathsf{P}\left(Z > \dfrac{0.81 - 0.8}{0.02}\right)$

$$= \mathsf{P}(Z > 0.5) = 0.5 - \mathsf{P}(0 < Z < 0.5) = 0.5 - 0.1915 = 0.3085$$

(b) $\mathsf{P}[(X > 0.825) \cup (X < 0.785)] = 2\mathsf{P}(X > 0.825)$

$$= 2\mathsf{P}\left(Z > \frac{0.025}{0.02}\right) = 2\mathsf{P}(Z > 1.25)$$

$$= 2[-\mathsf{P}(0 < Z < 1.25) + 0.5] = 2[-0.3944 + 0.5] = 0.2112$$

2. (a) $\sigma = 0.02,\qquad \mathsf{P}(X < 2) = 0.01$ We need to find μ from $\mathsf{P}\left(Z < \dfrac{2-\mu}{0.02}\right) = 0.01.$

$$\therefore \quad 0.05 - \mathsf{P}\left(0 < Z < \frac{\mu - 2}{0.02}\right) = 0.01 \quad \therefore \quad \frac{\mu - 2}{0.02} = 2.33 \quad \therefore \quad \mu = 2.0466$$

(b) Now we require σ such that $\mathsf{P}(X < 2) = 0.001$ with $\mu = 2.0466$

i.e. $0.5 - \mathsf{P}\left(0 < Z < \dfrac{0.0466}{\sigma}\right) = 0.001$

$$\therefore \quad \mathsf{P}\left(0 < Z < \frac{0.0466}{\sigma}\right) = 0.499 \quad \therefore \quad \frac{0.0466}{\sigma} = 3.1 \quad \therefore \quad \sigma = 0.015$$

3. $L \sim N(4.01,\ (0.03)^2)$

Cost has 2 possible values per usable rod: 6p, 8p.

$\mathsf{P}(C = 6) = \mathsf{P}(3.98 < L < 4.02) = \mathsf{P}\left(0 < Z < \dfrac{4.01 - 3.98}{0.03}\right) + \mathsf{P}\left(0 < Z < \dfrac{4.02 - 4.01}{0.03}\right)$

$$= \mathsf{P}(0 < Z < 1) + \mathsf{P}(0 < Z < 0.333) = 0.3413 + 0.1305 = 0.4718$$

$\mathsf{P}(C = 8) = \mathsf{P}(L > 4.02) = \mathsf{P}(Z > 0.333) = 0.5 - \mathsf{P}(0 < Z < 0.333) = 0.3695$

For every 100 rods produced:

	Total
36.95 are usable after shortening costing 8p each	295.6
47.18 are immediately usable costing 6p each	283.08
15.87 are scrap costing 5p each	79.35

Average cost per usable rod $= \dfrac{283.08 + 295.6 + 79.35}{84.13} = 7.82$

Answers

4. Let X = the amount of coffee in a fill; then $X \sim N(200, 7)$

   (a) $\mathrm{P}(X > 215) = \mathrm{P}(Z > \dfrac{215.0 - 200.0}{7.0}) = \mathrm{P}(Z > 2.14) = 0.016$ from Z-tables.

   Hence, from a consignment of 1000 packets, the number containing more than

   215 gm $= 1000 \times 0.016 = 16$

   (b) $\mathrm{P}(X < 195) = \mathrm{P}(Z < \dfrac{195.0 - 200.0}{7.0} = \mathrm{P}(Z < -0.714) = 0.2389$ from Z-tables.

   Hence, from a consignment of 1000 packets, the number containing less than

   195 gm $= 1000 \times 0.2389 = 238.9$

   (c)

$$\begin{aligned}\mathrm{P}(190.0 < X < 210.0) &= \mathrm{P}(\frac{190.0 - 200.0}{7.0} < Z < \frac{210.0 - 200.0}{7}) \\ &= \mathrm{P}(-1.43 < Z < 1.43) = 0.8472 \quad \text{from } Z\text{-tables.}\end{aligned}$$

   Hence, from a consignment of 1000 packets, the number containing between

   190 gm and 210 gm $= 1000 \times 0.8472 = 847$

   If 40 out of the 1000 packets are withdrawn, then P(sub-standard packet) $= \dfrac{40}{1000} = 0.04$.

   Let k be the limit below which packets are sub-standard, then $\mathrm{P}(X < k) = 0.04$

   From Z-tables, $Z = -1.75$ as we are dealing with 'less than' i.e. the 'left-hand' part of the standard normal distribution curve.

   Hence, $\dfrac{k - 200.0}{7} = -1.75$ i.e. $k = -1.75(7) + 200.0 = 187.75$

   'Line drawn' at 188 gm; any packet below this value to be withdrawn.

5. Let X be the time taken to assemble the component; then $X \sim N(110, 10)$

   (a) $\mathrm{P}(X < 95) = \mathrm{P}(Z < \dfrac{95.0 - 110.0}{10.0}) = \mathrm{P}(Z < -1.5) = 0.3085$ from Z-tables

   Hence, from a group of 20 teams, the number completing the assembly within

   95 minutes $= 20 \times 0.3085 = 6.17$ so the number of teams is 6.

   (b) $\mathrm{P}(X > 120) = \mathrm{P}(Z > \dfrac{120.0 - 110.0}{10.0}) = \mathrm{P}(Z > 1.0) = 0.1587$ from Z-tables

   Hence, from a group of 20 teams, the number completing the assembly in more than

   2 hours $= 20 \times 0.1587 = 3.174$ so the number of teams is 3.

   If 95% of teams are to complete the assembly 'on time', then 5% take longer than the set time, k, and $\mathrm{P}(X > k) = 0.05$ hence, $Z = 1.64$

   Therefore, $\dfrac{k - 110.0}{10.0} = 1.64$ or, $k = 10(1.64) + 110.0 = 126.4$ minutes.

Sums and Differences of Random Variables

39.3

Introduction

In some situations, it is possible to easily describe a problem in terms of sums and differences of random variables. Consider a typical situation in which shafts are fitted to cylindrical sleeves. One random variable is used to describe the variability of the diameter of the shaft, and one is used to describe the variability of the sleeves. Clearly, we need to know how the total variability involved affects the fitting of shafts and sleeves. In this Section, we will confine ourselves to cases where the random variables are normally distributed and independent.

Prerequisites

Before starting this Section you should ...

- be familiar with the results and concepts met in the study of probability
- be familiar with the normal distribution

Learning Outcomes

On completion you should be able to ...

- describe a variety of problems in terms of sums and differences of normal random variables
- solve problems described in terms of sums and differences of normal random variables

1. Sums and differences of random variables

In some situations, we can specify a problem in terms of sums and differences of random variables. Here we confine ourselves to cases where the random variables are normally distributed. Typical situations may be understood by considering the following problems.

Problem 1

In a certain mass-produced assembly, a 3 cm shaft must slide into a cylindrical sleeve. Shafts are manufactured whose diameter S follows a normal distribution $S \sim N(3, 0.004^2)$ and cylindrical sleeves are manufactured whose internal diameter C follows a normal distribution $C \sim N(3.010, 0.003^2)$. Assembly is performed by selecting a shaft and a cylindrical sleeve at random. In what proportion of cases will it be impossible to fit the selected shaft and cylindrical sleeve together?

Discussion

Clearly, the shaft and cylindrical sleeve will fit together only if the diameter of the shaft is smaller than the internal diameter of the cylindrical sleeve. We need the difference of the two random variables C and S to be greater than zero. We can take the difference $C - S$ and find its distribution. Once we do this we can then ask the question "What is the probability that the inside diameter of the cylindrical sleeve is greater than the outside diameter of the shaft, i.e. what is P$(C - S > 0)$?" Essentially we are trying to ensure that the internal diameter of the cylindrical sleeve is larger than the external diameter of the shaft.

Problem 2

A manufacturer produces boxes of woodscrews containing a variety of sizes for a local DIY store. The weight W (in kilograms) of boxes of woodscrews manufactured is a normal random variable following the distribution $W \sim N(1.01, 0.004)$. Note that 0.004 is the *variance*. Find the probability that a customer who selects two boxes of screws at random finds that their combined weight is greater than 2.03 kilograms.

Discussion

In this problem we are looking at the effects of adding two random variables together. Since all boxes are assumed to have weights W which follow the distribution $W \sim N(1.01, 0.004)$, we are considering the effect of adding the random variable W to itself. In general, there is no reason why we cannot combine variables $W_1 \sim N(\mu_1, \sigma_1^2)$ and $W_2 \sim N(\mu_2, \sigma_2^2)$. This might happen if the DIY store bought in two similar products from two different manufacturers.

Before we can solve such problems, we need to obtain some results concerning the behaviour of random variables.

Functions of several random variables

Note that we shall quote results only for the continuous case. The results for the discrete case are similar with integration replaced by summation. We will omit the mathematics leading to these results.

Key Point 4

- If $X_1, X_2, \cdots + X_n$ are n random variables then
$$\mathsf{E}(X_1 + X_2 + \cdots + X_n) = \mathsf{E}(X_1) + \mathsf{E}(X_2) + \ldots \mathsf{E}(X_n)$$
- If $X_1, X_2, \ldots X_n$ are n independent random variables then
$$\mathsf{V}(X_1 + X_2 + \cdots + X_n) = \mathsf{V}(X_1) + \mathsf{V}(X_2) + \cdots + \mathsf{V}(X_n)$$
and more generally
$$\mathsf{V}(X_1 \pm X_2 \pm \cdots \pm X_n) = \mathsf{V}(X_1) + \mathsf{V}(X_2) + \cdots + \mathsf{V}(X_n)$$

Example 16

Solve Problem 1 from the previous page. You may assume that the sum and difference of two normal random variables are themselves normal.

Solution

Consider the random variable $C - S$. Using the results above we know that

$$C - S \sim N(3.010 - 3.0, 0.004^2 + 0.003^2) \quad \text{i.e,} \quad C - S \sim N(0.01, 0.005^2)$$

Hence $\quad \mathsf{P}(C - S > 0) = \mathsf{P}(Z > \dfrac{0 - 0.01}{0.005} = -2) = 0.9772$

This result implies that in 2.28% of cases it will be impossible to fit the shaft to the sleeve.

Task

Solve Problem 2 from the previous page. You may assume that the sum and difference of two normal random variables are themselves normal.

Your solution

Answer

If W_{12} is the random variable representing the combined weight of the two boxes then

$$W_{12} \sim N(2.02, 0.008)$$

Hence

$$\mathsf{P}(W_{12} > 2.03) = \mathsf{P}(Z > \frac{2.03 - 2.02}{\sqrt{0.008}} = 0.1118) = 0.5 - 0.0445 = 0.4555$$

The result implies that the customer has about a 46% chance of finding that the weight of the two boxes combined is greater than 2.03 kilograms.

Exercises

1. Batteries of type A have mean voltage 6.0 (volts) and variance 0.0225 (volts^2). Type B batteries have mean voltage 12.0 and variance 0.04. If we form a series connection containing one of each type what is the probability that the combined voltage exceeds 17.4?

2. Nuts and bolts are made separately and paired at random. The nuts' diameters, in mm, are independently $N(10, 0.02)$ and the bolts' diameters, in mm, are independently $N(9.5, 0.02)$. Find the probability that a bolt is too large for its nut.

3. Certain cutting tools have lifetimes, in hours, which are independent and normally distributed with mean 300 and variance 10000.

   (a) Find the probability that

      (i) the total life of three tools is more than 1000 hours.

      (ii) the total life of four tools is more than 1000 hours.

   (b) In a factory each tool is replaced when it fails. Find the probability that exactly four tools are needed to accumulate 1000 hours of use.

   (c) Explain why the first sentence in this question can only be approximately, not exactly, true.

4. A firm produces articles whose length, X, in cm, is normally distributed with nominal mean $\mu = 4$ and variance $\sigma^2 = 0.1$. From time to time a check is made to see whether the value of μ has changed. A sample of ten articles is taken, the lengths are measured, the sample mean length $\bar{X}$ is calculated, and the process is adjusted if $\bar{X}$ lies outside the range (3.9, 4.1). Determine the probability, α, that the process is adjusted as a result of a sample taken when $\mu = 4$. Find the smallest sample size n which would make $\alpha \leq 0.05$.

Answers

1. $X_A \sim N(6, 0.0225) \qquad X_B \sim N(12, 0.04)$
   Series $X = X_A + X_B \sim N(18, 0.0625)$ as variances always add

$$\mathsf{P}\,(X > 17.4) = \mathsf{P}\left(Z > \frac{-0.6}{0.25}\right) = 0.5 + \mathsf{P}\left(0 < Z < \frac{0.6}{0.25}\right)$$

$$= 0.5 + \mathsf{P}(0 < Z < 2.4) = 0.5 + 0.4918 = 0.9918$$

2. Let the diameter of a nut be N. Let the diameter of a bolt be B. A bolt is too large for its nut if $N - B < 0$.

$$\begin{aligned}
\mathsf{E}(N-B) &= 10 - 9.5 = 0.5 \\
\mathsf{V}(N-B) &= 0.02 + 0.02 = 0.04 \\
N - B &\sim N(0.5, 0.04) \\
\mathsf{P}(N-B<0) &= \mathsf{P}\left(\frac{N-B-0.5}{0.2} < \frac{0-0.5}{0.2}\right) = \mathsf{P}(Z < -2.5) \\
&= \Phi(-2.5) = 1 - \Phi(2.5) = 1 - 0.99379 \\
&= 0.00621.
\end{aligned}$$

   The probability that a bolt is too large for its nut is 0.00621.

3. (a) Let the lifetime of tool i be T_i.

   (i)
$$\begin{aligned}
\mathsf{E}(T_1+T_2+T_3) &= 900 \\
\mathsf{V}(T_1+T_2+T_3) &= 30000 \\
(T_1+T_2+T_3) &\sim N(900, 30000) \\
\mathsf{P}(T_1+T_2+T_3 > 1000) &= \mathsf{P}\left(\frac{T_1+T_2+T_3-900}{\sqrt{30000}}\right) = \mathsf{P}(Z > 0.57735) \\
&= 1 - \Phi(0.57735) = 1 - 0.7181 = 0.2819
\end{aligned}$$

   (ii)
$$\begin{aligned}
\mathsf{E}(T_1+T_2+T_3+T_4) &= 1200 \\
\mathsf{V}(T_1+T_2+T_3+T_4) &= 40000 \\
(T_1+T_2+T_3+T_4) &\sim N(1200, 40000) \\
\mathsf{P}(T_1+T_2+T_3+T_4 > 1000) &= \mathsf{P}\left(\frac{T_1+T_2+T_3+T_4-1200}{\sqrt{40000}} > \frac{1000-1200}{\sqrt{40000}}\right) \\
&= \mathsf{P}(Z > -1) \\
&= 1 - \Phi(-1) = \Phi(1) = 0.8413
\end{aligned}$$

Answers

(b) Let the number of tools needed be N.

$$\begin{aligned}
\mathsf{P}(N \leq 3) &= \mathsf{P}(N=1) + \mathsf{P}(N=2) + \mathsf{P}(N=3) \\
&= \mathsf{P}(T_1 + T_2 + T_3 > 1000) = 0.2819 \\
\mathsf{P}(N \leq 4) &= \mathsf{P}(N=1) + \mathsf{P}(N=2) + \mathsf{P}(N=3) + \mathsf{P}(N=4) \\
&= \mathsf{P}(T_1 + T_2 + T_3 + T_4 > 1000) = 0.8413.
\end{aligned}$$

Hence $\mathsf{P}(N = 4) = \mathsf{P}(N \leq 4) - \mathsf{P}(N \leq 3) = 0.8413 - 0.2819 = 0.5594$.

(c) Lifetimes can not be negative. The normal distribution assigns non-zero probability density to negative values so it can only be an approximation in this case.

4.

$$\begin{aligned}
X &\sim N(4, 0.1) \\
X_1 + \cdots + X_{10} &\sim N(40, 1) \\
\bar{X} = (X_1 + \cdots + X_{10})/10 &\sim N(4, 0.01)
\end{aligned}$$

By symmetry $\mathsf{P}(\bar{X} < 3.9) = \mathsf{P}(\bar{X} > 4.1)$.

$$\begin{aligned}
\mathsf{P}(\bar{X} < 3.9) &= \mathsf{P}\left(\frac{\bar{X} - 4}{0.1} < \frac{3.9 - 4}{0.1}\right) = \mathsf{P}(Z < -1) \\
&= \Phi(-1) = 1 - \Phi(1) = 1 - 0.8413 \\
&= 0.1587
\end{aligned}$$

More generally

$$\begin{aligned}
X_1 + \cdots + X_n &\sim N(4n, 0.1n) \\
\bar{X} = (X_1 + \cdots + X_n)/n &\sim N(4, 0.1/n) \\
\mathsf{P}(\bar{X} < 3.9) &= \mathsf{P}\left(\frac{\bar{X} - 4}{\sqrt{0.1/n}} < \frac{3.9 - 4}{\sqrt{0.1/n}}\right) = \mathsf{P}\left(Z < \frac{-0.1}{\sqrt{0.1/n}} = -\sqrt{0.1n}\right) \\
&= \Phi(-\sqrt{0.1n}) = 1 - \Phi(\sqrt{0.1n}) \\
\alpha &= 2[1 - \Phi(\sqrt{0.1n})]
\end{aligned}$$

We require $\alpha \leq 0.05$.

$$\begin{aligned}
2[1 - \Phi(\sqrt{0.1n})] \leq 0.05 &\Leftrightarrow 1 - \Phi(\sqrt{0.1n}) \leq 0.025 \\
&\Leftrightarrow \Phi(\sqrt{0.1n}) \geq 0.975 \\
&\Leftrightarrow \sqrt{0.1n} \geq 1.96 \\
&\Leftrightarrow n \geq 10 \times 1.96^2 = 38.416
\end{aligned}$$

The smallest sample size which satisfies this is $n = 39$.

Table 1: The Standard Normal Probability Integral

$Z=\frac{x-\mu}{\sigma}$	0	1	2	3	4	5	6	7	8	9
0	0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
.1	0398	0438	0478	0517	0577	0596	0636	0675	0714	0753
.2	0793	0832	0871	0909	0948	0987	1026	1064	1103	1141
.3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
.4	1555	1591	1628	1664	1700	1736	1772	1808	1844	1879
.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
.6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
.7	2580	2611	2642	2673	2703	2734	2764	2794	2822	2852
.8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
.9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1.0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	3665	3686	3708	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3888	3907	3925	3944	3962	3980	3997	4015
1.3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1.5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1.6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2.0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2.1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2.2	4861	4865	4868	4871	4875	4878	4881	4884	4887	4890
2.3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2.4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2.5	4938	4940	4941	4943	4946	4947	4948	4949	4951	4952
2.6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2.7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2.8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2.9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986
	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9
	4987	4990	4993	4995	4997	4998	4998	4999	4999	4999