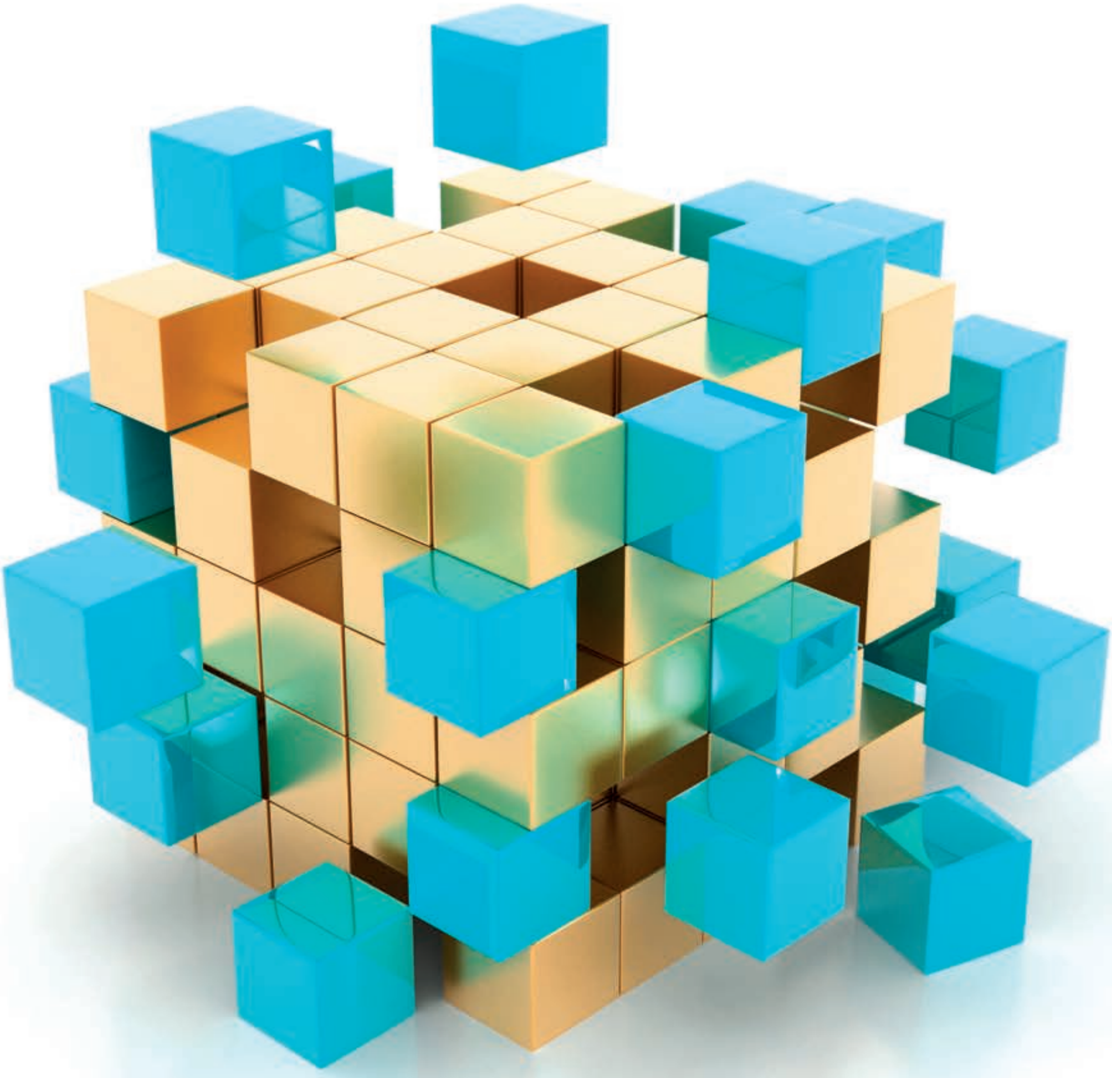
Quantitative Analysis for Decision Makers

Seventh edition

Mik Wisniewski & Farhad Shafti



Pearson

Seventh Edition

QUANTITATIVE ANALYSIS for Decision Makers

Mik Wisniewski

Freelance Consultant and Business Analyst

Farhad Shafti

Adam Smith Business School,
University of Glasgow

Pearson

Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • São Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

What is your conclusion? Effectively you have two choices: either the die is fixed or tampered with in some way or a highly unlikely sequence of events has actually happened. Given that the probability of rolling six sixes with a fair die is 0.00002 (the multiplication rule, remember?), it is tempting to conclude that the die has been tampered with in some way. However, based on this evidence alone we have no hard proof of this. It is a subjective assessment based on the (very low) probability of the event.

Let us return to our example. We are now in the position of having to assess this result and decide whether the actual defect rate is still 6 per cent or whether it might have worsened. We might argue that a 10 per cent probability is a reasonable one and that we cannot therefore say the defect rate has altered from its level of 6 per cent. On the other hand, we might argue that the probability of 10 per cent is too high to be likely to relate to the original actual 6 per cent defect rate. Like many situations we cannot prove one way or the other what has happened to the defect rate. Our assessment of the sample result must to some extent be subjective. As a compromise we might suggest that if the defect rate is of critical importance to the organisation – perhaps because of its quality policy or its costs – then we should repeat the sample and assess the second sample result. You should be able to see the logic of this. The only reason for rejecting the evidence of the first sample is to say that, by chance, we took more than a proportionate share of defective items from the assembly line. However, if we repeat the sample – particularly if we increase its size – then the chances are we should not repeat this 'bias': we would expect a new result closer to the three defective items we ought to be getting. If, on the other hand, we are still finding a higher number of defective items, this strengthens the case for concluding that the actual defect rate has worsened. A detailed investigation would then be needed to ascertain why this has happened and what corrective action could be taken.

Binomial tables

There is another method of obtaining Binomial probabilities: we can use pre-calculated tables and obtain the required probabilities directly. Then why, you ask, did we bother with the formula we have just introduced? The answer is that the tables typically only show the probabilities for certain combinations of n and p. If the required application falls outside these combinations we will need the formula in any event.

Typical tables are shown in Appendix A and we shall examine them shortly. To illustrate their use we shall introduce another typical situation. You may be aware that airlines operating passenger flights typically book more passengers onto a flight than there are seats available. They do this because experience has shown that not all passengers booked will actually turn up – they become 'no-shows'. Obviously, if the airline did not take this into account then the flight would typically depart with empty seats (and lost revenue). On the other hand, the airline does not wish to be in the embarrassing position of having too many passengers for a given flight (although this does happen). It is not only airlines who operate on this sort of basis: many organisations providing direct customer services operate on the same principle.

Assume the airline is operating a short-haul flight from Manchester to Edinburgh on a 16-seat aircraft. The typical passenger is an executive going to some business meeting and returning the same day. The airline knows from past experience that 15 per cent of passengers booked on their early morning flight will not appear. They have therefore adopted the practice of taking a maximum of 20 bookings for this flight. The customer service manager, however, is concerned about the likelihood of passengers booked on the flight not having a seat because more passengers than expected turn up for the flight. We have been asked to assess the probability of this happening. As usual we must make assumptions to apply the Binomial principles. We clearly have a sequence of 20 'trials': passengers booked on the flight. The probability of any one passenger not showing is 0.15, and to apply the Binomial we must make the assumption that the trials are independent. This implies that, for example, no two passengers are booked together (and hence will 'not-show' together). We then require the probability that there will be no more than three 'no-shows' (implying that we will have at least 17 passengers for 16 seats). Clearly this requires:

$$P(0 \text{ no shows}) + P(1 \text{ no shows}) + P(2 \text{ no shows}) + P(3 \text{ no shows})$$

That is, we will need several parts of the appropriate probability distribution. Obviously we could use the Binomial formula to work out each of these individual probabilities and the total (and you should do this later to practise the use of the formula approach). However, we can instead use the table in Appendix A, part of which is duplicated here for ease of reference.

Binomial distribution

n	r ≥	p = 0.15
20	0	1.0000
	1	0.9612
	2	0.8244
	3	0.5951
	4	0.3523
	5	0.1702
	6	0.0673
	7	0.0219
	8	0.0059
	9	0.0013
	10	0.0002

The table will require a little explanation. First, take note that it shows cumulative rather than individual probabilities, hence ≥. We shall explore this in a moment. Second, the table is segmented for various combinations of n, p and r, and we are interested in the combination of $n = 20$ and $p = 0.15$ with the values of r relating directly to our problem. The probabilities in the table are straightforward in their application. For example, we have:

$$P(r \geq 0) = 1.0000$$

$$P(r \geq 1) = 0.9612$$

$$P(r \geq 2) = 0.8244 \text{ and so on.}$$

In the context of our problem – where r effectively equates to a 'no-show' – we have a probability of 0.9612 that there will be at least one no-show, a probability of 0.8244 that there will be at least two and so on.

Progress Check 5.10

Using the table, determine the probability that more passengers arrive for the flight than there are seats available.

Applying some simple logic it is evident that directly from the table we can derive:

$$P(r \geq 4) = 0.3523$$

That is, there is a probability of 0.3523 that there will be at least four no-shows and, therefore, we will have sufficient seats for the remaining passengers. It follows then that we must have:

$$\begin{aligned} P(r < 4) &= P(r = 0) + P(r = 1) + P(r = 2) + P(r = 3) \\ &= 1 - 0.3523 = 0.6477 \end{aligned}$$

as the probability that there will not be enough no-shows to avoid an excess of passengers over seats. That is, a probability of 0.65 (approximately) that the airline will have overbooked. If the airline has not yet had complaints from irate passengers who have booked but couldn't get a seat, it soon will have!

Note also that, although the table relates to cumulative probabilities, it can be used to derive individual probabilities. Suppose we had actually wanted the probability of exactly three no-shows: $P(r = 3)$. From the table we have:

$$\begin{aligned} P(r \geq 4) &= P(r = 4) + P(r = 5) + \cdots + P(r = 20) = 0.3523 \\ P(r \geq 3) &= P(r = 3) + P(r = 4) + \cdots + P(r = 20) = 0.5951 \end{aligned}$$

Hence we must have:

$$\begin{aligned} P(r = 3) &= P(r \geq 3) - P(r \geq 4) \\ &= 0.5951 - 0.3523 = 0.2428 \end{aligned}$$

Mean and standard deviation of a Binomial distribution

It may have occurred to you that the Binomial distribution (or indeed any other probability distribution) is just a special example of distributions in general, like those we looked at in Chapter 4. For those distributions we were able to calculate a number of summary statistics such as the mean and standard deviation. It seems logical that we should be able to calculate comparable statistics for the Binomial distribution. In our airline example, the mean would indicate the mean number of no-shows on each flight, and the standard deviation would show the variability around this mean. We could calculate the mean and standard deviation using the formulae from Chapter 4 with the probabilities as the frequencies and the outcomes as the interval values. However, there is a more direct method. Without proof we state that for a Binomial distribution:

$$\text{Mean} = np$$

$$\text{Standard deviation} = \sqrt{npq}$$

Here we would have:

$$\text{Mean} = 20(0.15) = 3$$

$$\text{Standard deviation} = \sqrt{20(0.15)(0.85)} = 1.597$$

The mean and standard deviation for a Binomial problem can be interpreted and used as with any other set of data. We can expect a mean number of no-shows of three per flight with a standard deviation of 1.6 (the mean confirms that with a 16-seat aircraft available and 20 seats sold the airline is heading for trouble). We could also calculate the mean number of no-shows for a range of ticket sales (17, 18, 19, 20, 21, etc.) to assess the likely impact on overbooking.

Progress Check 5.11

You work for a mail-order retail company which advertises special promotions on the internet. Customers who respond to the promotion asking to buy the product are offered it on a sale-or-return basis. That is, the company sends the product to the customer together with an invoice. If the customer is happy with the product they will pay the invoice. If they are not happy, they can simply return the product to the company and do not have to pay the invoice. Obviously, in the latter case, the company has incurred costs which it cannot recoup (postage, handling, etc.). For past promotions the company has noted that 12 per cent of customers return the product. For the next promotion 50 000 customer orders are expected. Calculate the mean and standard deviation and comment on how this might be used by management.

Solution is given on p 577.

Binomial probabilities in Excel

Excel, and other spreadsheet packages, have built in statistical functions to calculate binomial probabilities directly. The function is:

BINOMDIST(r, n, p, cumulative)

Where:

r is the specified number of outcomes we wish to calculate for

n is the number of trials

p is the probability of the outcome in each trial

cumulative is a logical value that is set to be either TRUE or FALSE. If the logical value is set to TRUE then Excel calculates the cumulative probability up to and including r. If the logical value is set to FALSE then Excel calculates the probability of exactly r outcomes.

So, with the airline example from the previous section we had $n = 20$ and $p = 0.15$. Using the function as:

BINOMDIST(4, 20, 0, 15, TRUE)

then Excel would return a value of 0.6477 as the probability of $r \leq 4$. On the other hand, using the function as:

BINOMDIST(4, 20, 0, 15, FALSE)

Excel would return a value of 0.1821 as the probability of $r = 4$.

The normal distribution

We now turn to the second theoretical probability distribution – the *Normal distribution*. The Normal distribution is widely used in business and management decision making and underpins the area of statistical inference that we shall be examining in detail in Chapter 7. It is instantly recognisable graphically, as shown in Figure 5.3. The Normal distribution is symmetrical – often referred to as bell-shaped – and this general shape remains the same no matter what problem we are examining where the Normal distribution applies. This has important consequences, as we shall see shortly. What will vary from one application of this distribution to another is not the general shape but two key characteristics of that shape:

- the mean value;
- the variability as measured by the standard deviation.

Figure 5.4 illustrates two such distributions where the means vary and Figure 5.5 where the standard deviations vary.

We could build up an entire family of such distributions depending on the specific values of the mean and standard deviation. So how do we determine the appropriate probabilities for such a distribution? For the Binomial we saw that we could either apply a formula or we could use pre-calculated tables. In the case of the Normal distribution

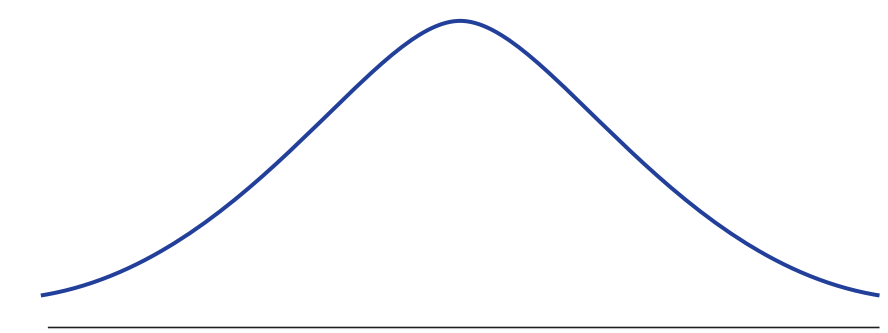


Figure 5.3 The Normal distribution

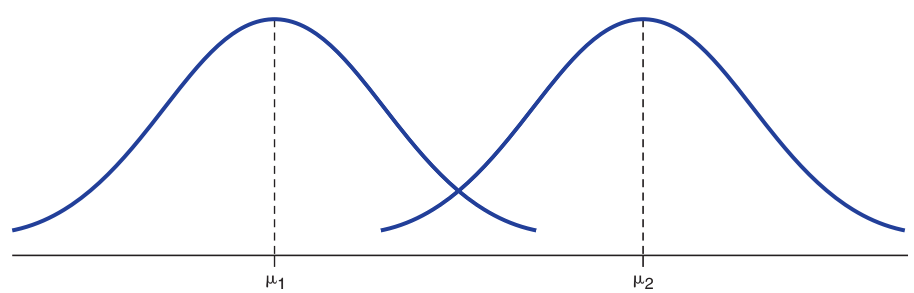


Figure 5.4 Two Normal distributions: differing means

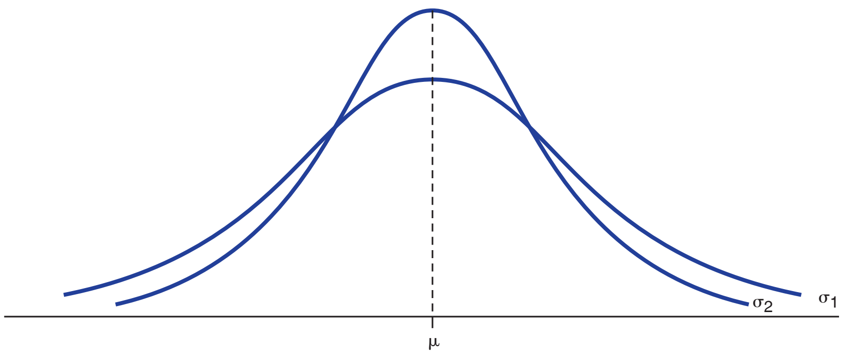


Figure 5.5 Two Normal distributions: differing standard deviations

the only feasible method is to use tables. The reason for this may become evident if we examine the formula behind the Normal distribution:

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

where:

μ is the mean

σ is the standard deviation

$\pi = 3.14159$

$e = 2.71828$

and x the value for which we require a probability.

Fortunately for all of us, tables of probabilities have been pre-calculated! However, before we can utilise them we must consider the issue of varying means and standard deviations in the family of Normal distributions. Remember for the Binomial distribution certain combinations of n and p gave rise to specific probability values. The number of combinations was fairly limited, partly because the Binomial deals with combinations of n and r which take only discrete values. We could always return to the Binomial formula if a particular combination did not appear in the tables. For the Normal distribution this problem is compounded. The number of possible combinations of mean and standard deviation we might require is much larger, since it is a continuous distribution, and if the particular combination we do require is not available the prospect of returning to the formula is not a happy one. However, we can take advantage of the fact that all Normal distributions follow the same general shape to use what is known as the *Standardised Normal distribution*.

The Standardised Normal distribution

Let us consider the following problem. A large retail organisation has recently been receiving a number of complaints from customers about one of its products: bottles of own-brand hair shampoo. Customers have complained that they think the actual contents of the shampoo bottles are sometimes less than they are meant to be according to