Chapter 7

Random Variables

Working with Uncertain Numbers

Chapter Outline

Many business situations involve random variables, such as waiting to find out your investment portfolio performance or asking customers in a marketing survey how much they would spend. Whenever a random experiment produces a number (or several numbers) as part of its outcome, you can be sure that random variables are involved. Naturally, you will want to be able to compute and interpret summary measures (such as typical value and risk) as well as probabilities of events that depend on the observed random quantity—for example, the probability that your portfolio grows by 10% or more.

You can also think about random variables as being where data sets come from. That is, many of the data sets you worked with in Chapters 2–5 were obtained as observations of random variables. In this sense, the random variable itself represents the population (or the process of sampling from the population), whereas the observed values of the random variable represent the sample data. Much more on population and samples is coming in Chapter 8 and beyond, but the fundamentals of random numbers are covered here in this chapter.

Here are some examples of random variables. Note that each one is random until its value is observed:

One: Next quarter's sales—a number that is currently unknown and that can take on one of a number of different values.
Two: The number of defective machines produced next week.
Three: The number of qualified people who will respond to your "help wanted" advertisement for a new full-time employee.
Four: The price per barrel of oil next year.
Five: The reported income of the next family to respond to your information poll.

A **random variable** may be defined as a specification or description of a numerical result from a random experiment. The value itself is called an **observation**. For example, "next quarter's sales" is a random variable because it specifies and describes the number that will be produced by the random experiment of waiting until next quarter's numbers are in and computing the sales. The actual future value, $3,955,846, is an observation of this random variable. Note the distinction between a random variable (which refers to the random process involved) and an observation (which is a fixed number, once it has been observed). The pattern of probabilities for a random variable is called its **probability distribution**.

Many random variables have a mean and a standard deviation.[1] In addition, there is a probability for each event based on a random variable. We will consider two

1. All of the random variables considered in this chapter have a mean and a standard deviation, although in theory there do exist random variables that have neither a mean nor a standard deviation.

Practical Business Statistics, Sixth Edition. DOI: 10.1016/B978-0-12-385208-3.00007-9

types of random variables: *discrete* and *continuous*. It is easier to work with a discrete random variable because you can make a list of all of its possible values. You will learn about two particular distributions that are especially useful: the *binomial distribution* (which is discrete) and the *normal distribution* (which is continuous). Furthermore, in many cases, you may use a (much simpler) normal probability calculation as a close approximation to a binomial probability.

Since there are so many different types of situations in which data values can arise, there are many types of random variables. The *exponential* and the *Poisson* distributions provide a look at the tip of this iceberg.

A random variable is **discrete** if you can list all possible values it can take on when it is observed. A random variable is **continuous** if it can take on any value in a range (for example, any positive number). For some random variables, it is unclear whether they are discrete or continuous. For example, next quarter's sales might be $385,298.61, or $385,298.62, or $385,298.63, or a similar amount up to some very large number such as $4,000,000.00. Literally speaking, this is discrete (since you can list all the possible outcomes). However, from a practical viewpoint, since the dividing points are so close together and no single outcome is very likely, you may work with it as if it were continuous.

7.1 DISCRETE RANDOM VARIABLES

When you have the list of values and probabilities (which defines the *probability distribution*) for a discrete random variable, you know everything possible about the process that will produce a random and uncertain number. Using this list, you will be able to calculate any summary measure (e.g., of typical value or of risk) or probability (of any event determined by the observed value) that might be of interest to you.

Here are some examples of discrete random variables:

1. The number of defective machines produced next week. The list of possible values is 0, 1, 2, ….
2. The number of qualified people who will respond to your "help wanted" advertisement for a new full-time employee. Again, the list of possible values is 0, 1, 2, ….
3. The resulting budget when a project is selected from four possibilities with costs of $26,000, $43,000, $54,000, and $83,000. The list of possible values is (in thousands of dollars) 26, 43, 54, and 83.

Such a list of possible values, together with the probability of each happening, is the probability distribution of the discrete random variable. These probabilities must be positive numbers (or 0) and must add up to 1. From this distribution, you can find the mean and standard deviation of the random variable. You can also find the probability of any event, simply by adding up the probabilities in the table that correspond to the event.

Example

Profit under Various Economic Scenarios

During a brainstorming session devoted to evaluation of your firm's future prospects, there was a general discussion of what might happen in the future. It was agreed to simplify the situation by considering a best-case scenario, a worst-case scenario, and two intermediate possibilities. For each of these four scenarios, after considerable discussion, there was general agreement on the approximate profit that might occur and its likelihood. Note that this defines the *probability distribution* for the random variable "profits" because we have a list of values and probabilities: one column shows the values (in this case, profit) and another column shows the probabilities.

Economic Scenario	Profit ($ millions)	Probability
Great	10	0.20
Good	5	0.40
OK	1	0.25
Lousy	–4	0.15

This probability distribution can be easily used to find probabilities of all events concerning profit. The probability that the profit is $10 million, for example, is 0.20. The probability of making $3 million or more is found as follows: $0.20 + 0.40 = 0.60$—because there are two outcomes ("Great" and "Good") that correspond to this event by having profit of $3 million or more.

Finding the Mean and Standard Deviation

The **mean or expected value** of a discrete random variable is an exact number that summarizes it in terms of a typical value, in much the same way that the average summarizes a list of data.[2] The mean is denoted by the lowercase Greek letter μ (*mu*) or by $E(X)$ (read as "expected value of X") for a random variable X. The formula is

Mean or Expected Value of a Discrete Random Variable *X*

$$\begin{aligned}\mu = E(X) &= \text{Sum of (value times probability)}\\ &= \sum XP(X)\end{aligned}$$

If the probabilities were all equal, this would be the average of the values. In general, the mean of a random variable is a weighted average of the values using the probabilities as weights.

This mean profit in the preceding example is

$$\begin{aligned}\text{Expected profit} &= (10\times 0.20) + (5\times 0.40) + (1\times 0.25)\\ &\quad + (-4\times 0.15) = 3.65\end{aligned}$$

2. In fact, the mean of a random variable is also called its average; however, we will often use *mean* for random variables and *average* for data.

Thus, the expected profit is \$3.65 million. This number summarizes the various possible outcomes (10, 5, 1, –4) using a single number that reflects their likelihoods.

The **standard deviation** of a discrete random variable indicates approximately how far you expect it to be from its mean. In many business situations, the standard deviation indicates the *risk* by showing just how uncertain the situation is. The standard deviation is denoted by σ, which matches our use of σ as the population standard deviation. The formula is

Standard Deviation of a Discrete Random Variable *X*

$$\sigma = \sqrt{\text{Sum of (squared deviation times probability)}}$$
$$= \sqrt{\sum (X-\mu)^2 P(X)}$$

Note that you would *not* get the correct answer by simply using the Σ key on your calculator to accumulate only the single column of values, since this would not make proper use of the probabilities.

The standard deviation of profit for our example is

$$\begin{aligned}\sigma &= \sqrt{\{[(10-3.65)^2 0.20] + [(5-3.65)^2 0.40] + [(1-3.65)^2 0.25] + [(-4-3.65)^2 0.15]\}}\\ &= \sqrt{8.064500 + 0.729000 + 1.755625 + 8.778375}\\ &= \sqrt{19.3275} = 4.40\end{aligned}$$

The standard deviation of \$4.40 million shows that there is considerable risk involved here. Profit might reasonably be about \$4.40 million above or below its mean value of \$3.65 million. Table 7.1.1 shows the details of the computations involved in finding the standard deviation.

To use Excel® to compute the mean and standard deviation of a discrete random variable, you might proceed as follows. Using Excel's menu commands, give names to these columns by selecting the numbers with the titles, then choosing Excel's Create from Selection in the Defined names section of the Formulas Ribbon. The mean (3.65) is the sum of the products of value times probability; hence, the formula is "=SUMPRODUCT (Profit,Probability)." Give this cell (which now contains the mean) the name "Mean." The standard deviation (4.40) is the square root (SQRT) of the sum of the products of the square of value minus mean times probability. Hence, the formula is

$$= \text{SQRT(SUMPRODUCT((Profit} - \text{Mean)\^{}2, Probability))}$$

These formulas give us 3.65 for the mean and 4.40 for the standard deviation, as before.

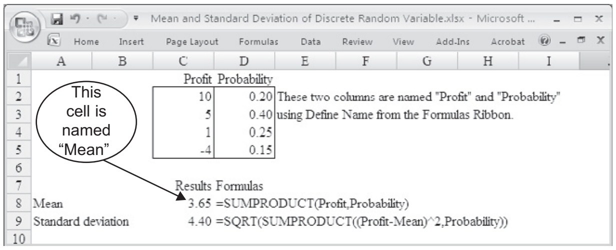


Figure 7.1.1 shows the probability distribution, with the heights of the lines indicating the probability and the location of the lines indicating the amount of profit in each case. Also indicated is the expected value, \$3.65 million, and the standard deviation, \$4.40 million.

Example
Evaluating Risk and Return

Your job is to evaluate three different projects (*X*, *Y*, and *Z*) and make a recommendation to upper management. Each project requires an investment of \$12,000 and pays off next year. Project *X* pays \$14,000 for sure. Project *Y* pays either \$10,000 or \$20,000 with probability 0.5 in each case. Project *Z* pays nothing with probability 0.98 and \$1,000,000 with probability 0.02. A summary is shown in Table 7.1.2.

(*Continued*)

TABLE 7.1.1 Finding the Standard Deviation for a Discrete Random Variable

Profit	Probability	Deviation from Mean	Squared Deviation	Squared Deviation Times Probability
10	0.20	6.35	40.3225	8.064500
5	0.40	1.35	1.8225	0.729000
1	0.25	–2.65	7.0225	1.755625
–4	0.15	–7.65	58.5225	8.778375
				Sum: 19.3275
				Square root: 4.40

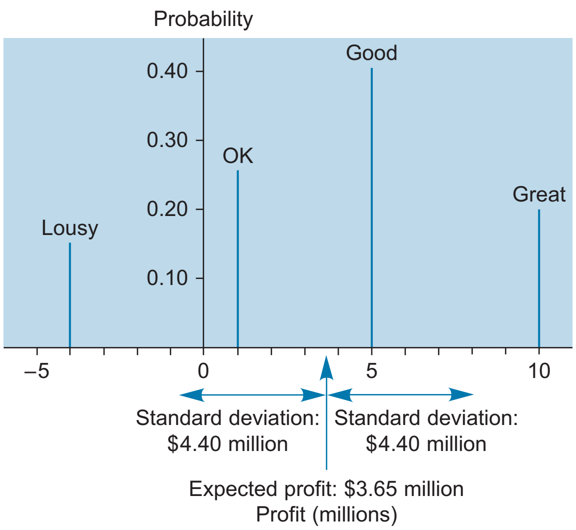


FIGURE 7.1.1 The probability distribution of future profits, with the mean (expected profits) and standard deviation (risk) indicated.

TABLE 7.1.2 Payoffs and Probabilities for Three Projects

Project	Payoff	Probability
X	14,000	1.00
Y	10,000	0.50
	20,000	0.50
Z	0	0.98
	1,000,000	0.02

Example—cont'd

The means are easily found: $14,000 for *X*, $10{,}000 \times 0.50 + 20{,}000 \times 0.50 = \$15{,}000$ for *Y*, and $0 \times 0.98 + 1{,}000{,}000 \times 0.02 = \$20{,}000$ for *Z*. We could write these as follows:

$$E(X) = \mu_X = \$14{,}000$$
$$E(Y) = \mu_Y = \$15{,}000$$
$$E(Z) = \mu_Z = \$20{,}000$$

Based only on these expected values, it would appear that *Z* is best and *X* is worst. However, these mean values don't tell the whole story. For example, although project *Z* has the highest expected payoff, it also involves considerable risk: 98% of the time there would be no payoff at all! The risks involved here are summarized by the standard deviations:

$$\sigma_X = \sqrt{(14{,}000 - 14{,}000)^2 \times 1.00} = \$0$$

$$\sigma_Y = \sqrt{(10{,}000 - 15{,}000)^2 \times 0.50 + (20{,}000 - 15{,}000)^2 \times 0.50}$$
$$= \$5{,}000$$

$$\sigma_Z = \sqrt{(0 - 20{,}000)^2 \times 0.98 + (1{,}000{,}000 - 20{,}000)^2 \times 0.02}$$
$$= \$140{,}000$$

These standard deviations confirm your suspicions. Project *Z* is indeed the riskiest—far more so than either of the others. Project *X* is the safest—a sure thing with no risk at all. Project *Y* involves a risk of $5,000.

Which project should be chosen? This question cannot be answered by statistical analysis alone. Although the expected value and the standard deviation provide helpful summaries to guide you in choosing a project, they do not finish the task. Generally, people prefer larger expected payoffs and lower risk. However, with the choices presented here, to achieve a larger expected payoff, you must take a greater risk. The ultimate choice of project will involve your (and your firm's) "risk versus return" preference to determine whether or not the increased expected payoff justifies the increased risk.[3]

What if you measure projects in terms of profit instead of payoff? Since each project involves an initial investment of $12,000, you can convert from payoff to profit by subtracting $12,000 from each payoff value in the probability distribution table:

$$\text{Profit} = \text{Payoff} - \$12{,}000$$

Using the rules from Section 5.4, which apply to summaries of random variables as well as to data, subtract $12,000 from each mean value and leave the standard deviation alone. Thus, without doing any detailed calculations, you come up with the following expected profits:

X:	$2,000
Y:	$3,000
Z:	$8,000

The standard deviations of profits are the same as for payoffs, namely:

X:	$0
Y:	$5,000
Z:	$140,000

3. In your finance courses, you may learn about another factor that is often used in valuing projects, namely, the correlation (if any) between the random payoffs and the payoffs of a market portfolio. This helps measure the *diversifiable* and *nondiversifiable risk* of a project. Correlation (a statistical measure of association) will be presented in Chapter 11. The nondiversifiable component of risk is also known as systematic or systemic risk because it is part of the entire economic system and cannot be diversified away.

7.2 THE BINOMIAL DISTRIBUTION

Percentages play a key role in business. When a percentage is arrived at by counting the number of times something happens out of the total number of possibilities, the number of occurrences might follow a *binomial distribution*. If so, there are a number of time-saving shortcuts available for finding the expected value, standard deviation, and probabilities of various events. Sometimes you will be interested in the percentage; at other times the number of occurrences will be more relevant. The binomial distribution can give answers in either case. Here are some examples of random variables that follow a binomial distribution:

1. The number of orders placed, out of the next three telephone calls to your catalog order desk.
2. The number of defective products out of 10 items produced.
3. The number of people who said they would buy your product, out of 200 interviewed.
4. The number of stocks that went up yesterday, out of all issues traded on major exchanges.
5. The number of female employees in a division of 75 people.
6. The number of Republican (or Democratic) votes cast in the next election.

Definition of Binomial Distribution and Proportion

Focus attention on a particular event. Each time the random experiment is run, either the event happens or it doesn't. These *two* possible outcomes give us the *bi* in *binomial*. A random variable X, defined as the *number of occurrences* of a particular event out of n trials, has a **binomial distribution** if

1. For each of the n trials, the event always has the same probability π of happening.
2. The trials are independent of one another.

The independence requirement rules out "peeking," as in the case of the distribution of people who order the special at a restaurant. If some people order the special because they see other customers obviously enjoying the rich, delicious combination of special aromatic ingredients, and say, "WOW! I'll have that too!" the number who order the special would *not* follow a binomial distribution. Choices have to be made independently in order to get a binomial distribution.

The **binomial proportion** p is the binomial random variable X expressed as a fraction of n:

Binomial Proportion

$$p = \frac{X}{n} = \frac{\text{Number of occurrences}}{\text{Number of trials}}$$

(Note that π is a fixed number, the probability of occurrence, whereas p is a random quantity based on the data.) For example, if you interviewed $n = 600$ shoppers and found that $X = 38$ plan to buy your product, then the binomial proportion would be

$$p = \frac{X}{n} = \frac{38}{600} = 0.063, \text{ or } 6.3\%$$

The binomial proportion p is also called a *binomial fraction*. You may have recognized it as a relative frequency, which was defined in Chapter 6.

Example

How Many Orders Are Placed? The Hard Way to Compute

This example shows the hard way to analyze a binomial random variable. Although it is rarely necessary to draw the probability tree, since it is usually quite large, seeing it once will help you understand what is really going on with the binomial distribution. Furthermore, when the shortcut computations are presented (the easy way) you will appreciate the time they save!

Suppose you are interested in the next $n = 3$ telephone calls to the catalog order desk, and you know from experience (or are willing to assume[4]) that $\pi = 0.6$, so that 60% of calls will result in an order (the others are primarily calls for information, or misdirected). What can we say about the number of calls that will result in an order? Certainly, this number will be either 0, 1, 2, or 3 calls. Since a call is more likely to result in an order than not, we should probably expect the probability of getting three orders to be larger than the probability of getting none at all. But how can we find these probabilities? The probability tree provides a complete analysis, as shown in Figure 7.2.1a, indicating the result of each of the three phone calls.

Note that the conditional probabilities along the branches are always 0.60 and 0.40 (the individual probabilities for each call) since we assume orders occur independently and do not influence each other. The number of orders is listed at the far right in Figure 7.2.1a; for example, the second number from the top, 2, reports the fact that the first and second (but not the third) callers placed an order, resulting in two orders placed. Note that there are three ways in which two orders could be placed. To construct the probability distribution of the number of orders placed, you could add up the probabilities for the different ways that each number could happen:

Number of Callers Who Ordered, X	Percentage Who Ordered, $p = X/n$	Probability
0	0.0%	0.064
1	33.3%	0.288 (= 0.096 + 0.096 + 0.096)
2	66.7%	0.432 (= 0.144 + 0.144 + 0.144)
3	100.0%	0.216

This probability distribution is displayed in Figure 7.2.1b.

Now that you have the probability distribution, you can find all of the probabilities by adding the appropriate ones. For example, the probability of at least two orders is 0.432 + 0.216 = 0.648. You can also use the formulas for the mean and standard deviation from Section 7.1 to find the mean value (1.80 orders) and the standard deviation (0.849 orders). However, *this would be too much work!* There is a much quicker formula for finding the mean, standard deviation,

(*Continued*)

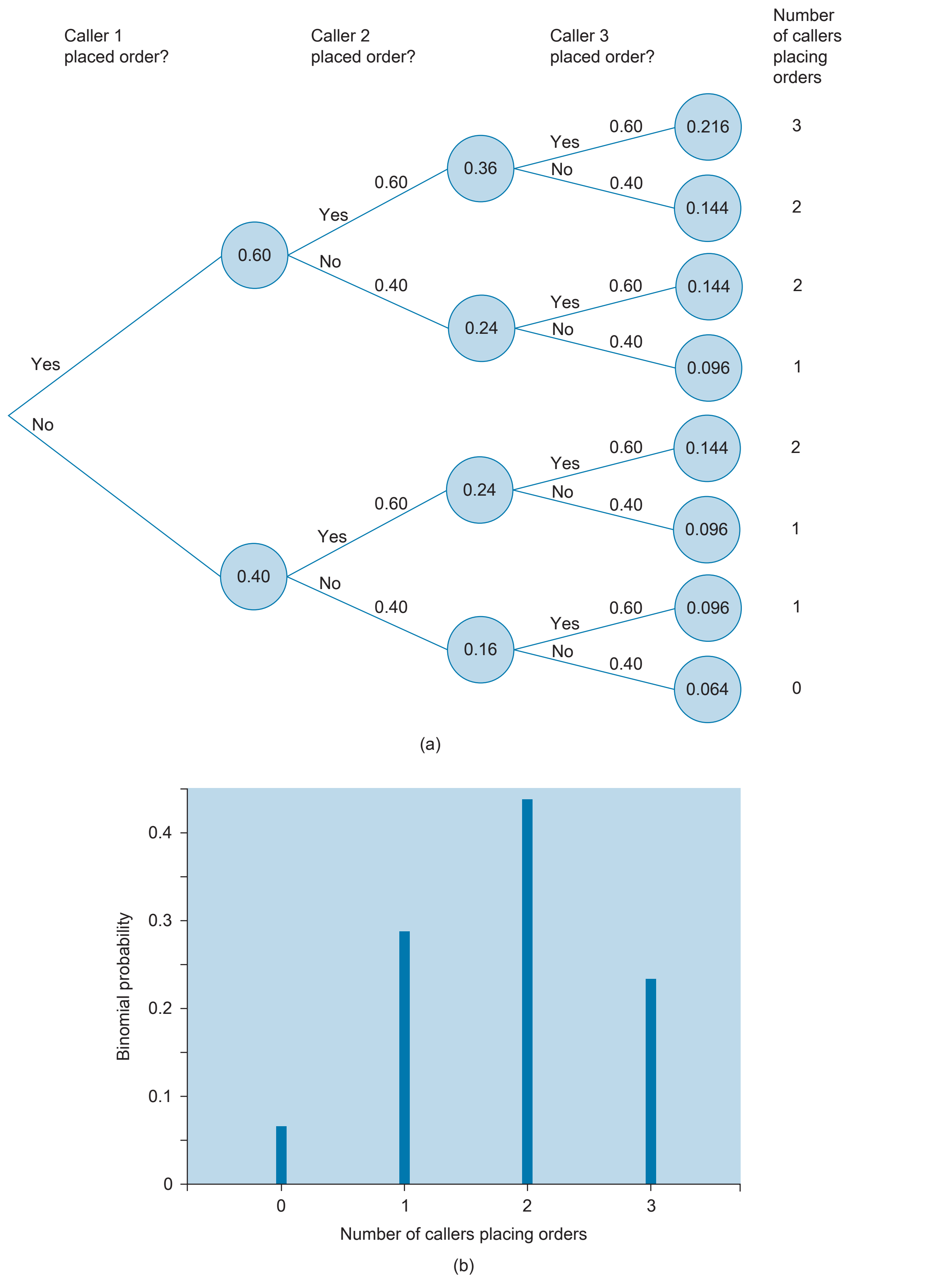


FIGURE 7.2.1 a. The probability tree for three successive telephone calls, each of which either does or does not result in an order being placed. There are eight combinations (the circles at the far right). In particular, there are three ways in which exactly two calls could result in an order: the second, third, and fifth circles from the top, giving a probability of $3 \times 0.144 = 0.432$. b. The binomial probability distribution of the number of calls that result in an order being placed.

Example—cont'd

and probabilities. Although it was possible to compute directly in this small example, you will not usually be so lucky. For example, had you considered 10 successive calls instead of 3, there would have been 1,024 probabilities at the right of the probability tree instead of the 8 in Figure 7.2.1.

4. The probability π is usually given in textbook problems involving a binomial distribution. In real life, they arise just as other probabilities do: from relative frequency, theoretical probability, or subjective probability.

Think of this example as a way of seeing the underlying situation and all combinations and then simplifying to a probability distribution of the *number of occurrences*. Conceptually, this is the right way to view the situation. Now let's learn the easy way to compute the answers.

Finding the Mean and Standard Deviation the Easy Way

The mean number of occurrences in a binomial situation is $E(X) = n\pi$, the number of possibilities times the probability of occurrence. The mean proportion is

$$E\left(\frac{X}{n}\right) = E(p) = \pi$$

which is the same as the individual probability of occurrence.[5]

This is what you would expect. For example, in a poll of a sample of 200 voters, if each has a 58% chance of being in favor of your candidate, on average, you would expect that

$$E\left(\frac{X}{n}\right) = E(p) = \pi = 0.58$$

or 58% of the sample will be in your favor. In terms of the number of people, you would expect $E(X) = n\pi = 200 \times 0.58 = 116$ people out of the 200 in the sample to be in your favor. Of course, the actually observed number and percentage will probably randomly differ from these expected values.

There are formulas for the standard deviation of the binomial number and percentage, summarized along with the expected values in the following table:

5. You might have recognized X/n as the relative frequency of the event. The fact that $E(X/n)$ is equal to π says that, on average, the relative frequency of an event is equal to its probability. In Chapter 8 we will learn that this property says that p is an unbiased estimator of π.

Mean and Standard Deviation for a Binomial Distribution

	Number of Occurrences, X	Proportion or Percentage, $p = X/n$
Mean	$E(X) = \mu_X = n\pi$	$E(p) = \mu_p = \pi$
Standard deviation	$\sigma_X = \sqrt{n\pi(1-\pi)}$	$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$

For the "telephone orders" example, we have $n = 3$ and $p = 0.60$. Using the formulas, the mean and standard deviation are

	Number of Occurrences, X	Proportion or Percentage, $p = X/n$
Mean	$E(X) = n\pi$ $= 3 \times 0.60$ $= 1.80$ calls	$E(X) = \pi$ $= 0.60$ or 60%
Standard deviation	$\sigma_X = \sqrt{n\pi(1-\pi)}$ $= \sqrt{3 \times 0.60(1-0.60)}$ $= 0.849$ calls	$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$ $= \sqrt{\frac{0.60(1-0.60)}{3}}$ $= 0.283$ or 28.3%

Thus, we expect 1.80 of these 3 telephone calls to result in an order. Sometimes more (i.e., 2 or 3) and sometimes fewer (i.e., 0 or 1) calls will result in an order. The extent of this uncertainty is measured (as usual) by the standard deviation, 0.849 calls. Similarly, we expect 60% of these 3 calls to result in an order. The last number, 28.3%, representing the standard deviation of the percentage, is interpreted as *percentage points* rather than as a percentage of some number. That is, while the expected percentage is 60%, the actual observed percentage is typically about 28.3 percentage points above this value (at 60 + 28.3 = 88.3%) or below (at 60 – 28.3 = 31.7%). This is natural if you remember that a standard deviation is stated in the same units as the data, which are percentage points in the case of p.

Example

Recalling Advertisements

Your company is negotiating with a marketing research firm to provide information on how your advertisements are doing with the American consumer. Selected people are to come in one day to watch TV programs and ads (for many products from many companies) and return the next day to answer questions. In particular, you plan to measure the *rate of recall*, which is the percentage of people who remember your ad the day after seeing it.

Before you contract with the firm to do the work, you are curious about how reliable and accurate the results

(*Continued*)

Example—cont'd

are likely to be. Your budget allows 50 people to be tested. From your discussions with the research firm, it seems reasonable initially to assume that 35% of people will recall the ad, although you really don't know the exact proportion. Based on the assumption that it really is 35%, how accurate will the results be? That is, about how far will the measured recall percentage be from the assumed value $\pi = 0.35$ with $n = 50$ for a binomial distribution? The answer is

$$\begin{aligned}\sigma_p &= \sqrt{\frac{\pi(1-\pi)}{n}}\\ &= \sqrt{\frac{0.35(1-0.35)}{50}}\\ &= 0.0675 \text{ or } 6.75\%\end{aligned}$$

This says that the standard deviation of the result of the recall test (namely, the percentage of people tested who remembered the ad) is likely to differ from the true percentage for the entire population typically by about 7 percentage points in either direction (above or below).

You decide that the results need to be more precise than that. The way to improve the precision of the results is to gather more information by increasing the sample size, n. Checking the budget and negotiating over the rates, you find that $n = 150$ is a possibility. With this larger sample, the standard deviation decreases to reflect the extra information:

$$\begin{aligned}\sigma_p &= \sqrt{\frac{\pi(1-\pi)}{n}}\\ &= \sqrt{\frac{0.35(1-0.35)}{150}}\\ &= 0.0389 \text{ or } 3.89\%\end{aligned}$$

You are disappointed that the extra cost didn't bring a greater improvement in the results. When the size of the study was tripled, the precision didn't even double! This is due, technically, to the fact that it is the *square root of n*, rather than n itself, that is involved. Nevertheless, you decide that the extra accuracy is worth the cost.

Finding the Probabilities

Suppose you have a binomial distribution, you know the values of n and π, and you want to know the probability that X will be exactly equal to some number a. There is a formula for this probability that is useful for small to moderate n. (When n is large, an approximation based on the normal distribution, to be covered in Section 7.4, will be much easier than the exact method presented here.) In addition, Table D–3 in Appendix D gives exact binomial probabilities and cumulative probabilities for $n = 1$ to 20 and $\pi = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$, and 0.95. Here is the exact formula:[6]

Binomial Probability That X Equals a

$$\begin{aligned}P(X=a) &= \binom{n}{a}\pi^a(1-\pi)^{n-a}\\ &= \frac{n!}{a!(n-a)!}\pi^a(1-\pi)^{n-a}\\ &= \frac{1\times2\times3\times\cdots\times n}{(1\times2\times3\times\cdots\times a)[1\times2\times3\times\cdots\times(n-a)]}\pi^a(1-\pi)^{n-a}\end{aligned}$$

By using this formula with each value of a from 0 to n (sometimes a lot of work), you (or a computer) can generate the entire probability distribution. From these values, you can find any probability you want involving X by adding together the appropriate probabilities from this formula.

To see how to use the formula, suppose there are $n = 5$ possibilities with a success probability $p = 0.8$ for each one, and you want to find the probability of exactly $a = 3$ successes. The answer is

$$\begin{aligned}P(X=3) &= \binom{5}{3}0.8^3(1-0.8)^{5-3}\\ &= \frac{5!}{3!(5-3)!}0.8^3\times0.2^2\\ &= \frac{1\times2\times3\times4\times5}{(1\times2\times3)(1\times2)}0.512\times0.040\\ &= 10\times0.02048 = 0.2048\end{aligned}$$

6. The notation $n!$ is read as "n factorial" and is the product of the numbers from 1 to n. For example, $4! = 1\times2\times3\times4 = 24$. (By convention, to get the correct answers, we define $0!$ to be 1.) Many calculators have a factorial key that works for values of n from 0 through 69. The notation

$$\binom{n}{a} = \frac{n!}{a!(n-a)!}$$

is the *binomial coefficient*, read aloud as "n choose a," and also represents the number of *combinations* you can make by choosing a items from n items (where the order of selection does not matter). Thus, it represents the number of different ways in which you could assign exactly a occurrences to the n possibilities. For example, with $n = 5$ and $a = 3$, the binomial coefficient is

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{120}{6\times2} = 10$$

Thus, there are 10 different ways (combinations) in which three out of five people could buy our product: the first three people could, or the first two and the fourth might, and so forth. The full list of the 10 combinations is (1,2,3), (1,2,4), (1,2,5), (1,3,4), (1,3,5), (1,4,5), (2,3,4), (2,3,5), (2,4,5), and (3,4,5).

This is the probability of *exactly* three successes. If you want the probability of *three or more* successes, you could compute the formula twice more: once for $a = 4$ and once for $a = 5$; the probability of three or more successes would be the total of these numbers. Alternatively, you could use a computer to obtain the probabilities, for example:

Probability Density Function and Cumulative Distribution Function

Binomial with $n = 5$ and $p = 0.800000$

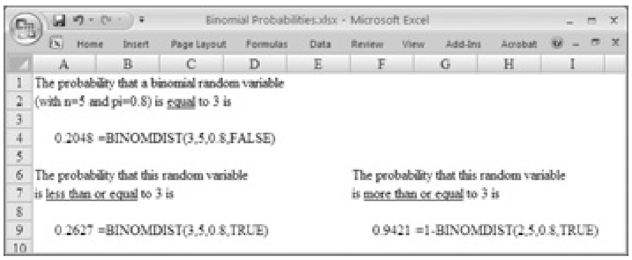
a	$P(X = a)$	$P(X <= a)$
0	0.0003	0.0003
1	0.0064	0.0067
2	0.0512	0.0579
3	0.2048	0.2627
4	0.4096	0.6723
5	0.3277	1.0000

In either case, once you have the individual probabilities (for 3, 4, and 5 successes), the answer is

$$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5)$$
$$= 0.2048 + 0.4096 + 0.3277$$
$$= 0.9421$$

Thus, you have a 94.2% chance of achieving three or more successes out of these five. Alternatively, using the complement rule, the probability of *three or more* must be one minus the probability of *two or less*, which is listed as 0.0579 in the computer output. The answer would then be found as 1 – 0.0579 = 0.9421.

To use Excel® to compute binomial probabilities, use the formula "=BINOMDIST(a,n,π, FALSE)" to find the probability $P(X = a)$ of being equal to a, and use the formula "=BINOMDIST(a,n,π,TRUE)" to find the probability $P(X \leq a)$ of being *less than or equal* to a, as follows:[7]



7. The "FALSE" and "TRUE" in Excel's binomial distribution formula refer to whether or not the probability distribution is cumulative, i.e., whether or not it accumulates probabilities for all of the previous (smaller) values of a as well.

Example

How Many Major Customers Will Call Tomorrow?

How many of your $n = 6$ major customers will call tomorrow? You are willing to assume that each one has a probability $\pi = 0.25$ of calling and that they call independently of one another. Thus, the number of major customers that will call tomorrow, X, follows a binomial distribution.

How many do you expect will call? That is, what is the expected value of X? The answer is $E(X) = n \times \pi = 1.5$ major customers. The standard deviation is $\sigma_X = \sqrt{6 \times 0.25 \times (1 - 0.25)} = 1.060660$, indicating that you can reasonably anticipate 1 or 2 more or less than the 1.5 you expect. Although this gives you an idea of what to expect, it doesn't tell you the chances that a given number will call. Let's compute the probabilities for this.

What is the probability that exactly $a = 2$ out of your $n = 6$ major customers will call? The answer is

$$P(X = 2) = \binom{6}{2} 0.25^2 (1 - 0.25)^{(6-2)}$$
$$= 15 \times 0.0625 \times 0.316406 = 0.297$$

Here is the entire probability distribution of the number of major customers who will call you tomorrow, including all possibilities for the number a from 0 through $n = 6$:

Probability Density Function and Cumulative Distribution Function

Binomial with $n = 6$ and $p = 0.250000$

a	$P(X = a)$	$P(X <= a)$
0	0.1780	0.1780
1	0.3560	0.5339
2	0.2966	0.8306
3	0.1318	0.9624
4	0.0330	0.9954
5	0.0044	0.9998
6	0.0002	1.0000

Note that the most likely outcomes are 1 or 2 calls, just as you suspected based on the mean value of 1.5 calls.

From this probability distribution, you can compute any probability about the number of major customers who will call you tomorrow. It is highly unlikely that all 6 will call (0.0002 or 0.02%, much less than a 1% chance). The probability that 4 or more will call is 0.0330 + 0.0044 + 0.0002 = 0.0376. From the second column, you can see that the probability that 3 or fewer will call is 0.9624. Your chances of spending a quiet day with no calls is 0.178. This probability distribution is shown in Figure 7.2.2.

Example

How Many Logic Analyzers to Schedule for Manufacturing?

You pay close attention to quality in your production facilities, but the logic analyzers you make are so complex that

(*Continued*)

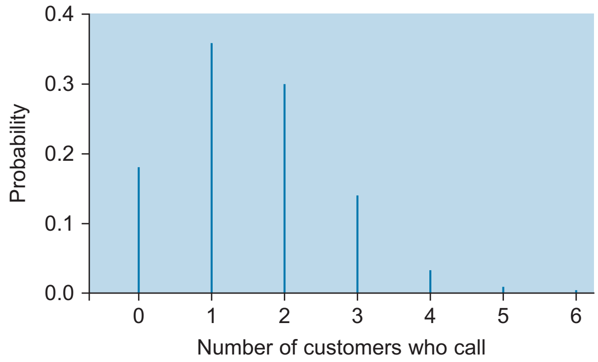


FIGURE 7.2.2 The probability distribution of the number of major customers who will call you tomorrow. These are binomial probabilities, with each vertical bar found using the formula based on $n = 6$ and $p = 0.25$. The number a is found along the horizontal axis.

Example—cont'd

there are still some failures. In fact, based on past experience, about 97% of the finished products are in good working order. Today you will have to ship 17 of these machines. The question is: How many should you schedule for production to be reasonably certain that 17 working logic analyzers will be shipped?

It is reasonable to assume a binomial distribution for the number of working machines produced, with n being the number that you schedule and π being each one's probability (0.97) of working. Then you can compute the probability that 17 or more of the scheduled machines will work.

What happens if you schedule 17 machines, with no margin for error? You might think that the high (97%) rate would help you, but, in fact, the probability that all 17 machines will work (using $n = 17$ and $a = 17$) is just 0.596:

$$\begin{aligned} P(X = 17 \text{ working machines}) &= \binom{17}{17} 0.97^{17} 0.03^{0} \\ &= 1 \times 0.595826 \times 1 = 0.596 \end{aligned}$$

Thus, if you schedule the same number, 17, that you need to ship, you will be taking a big chance! There is only a 59.6% chance that you will meet the order, and a 40.4% chance that you will fail to ship the entire order in working condition. The probability distribution is shown in Figure 7.2.3.

It looks as though you'd better schedule more than 17. What if you schedule $n = 18$ units for production? To find the probability that *at least* 17 working analyzers will be shipped, you'll need to find the probabilities for $a = 17$ and $a = 18$ and add them up:

$$\begin{aligned} P(X \geq 17) &= P(X = 17) + P(X = 18) \\ &= \binom{18}{17} 0.97^{17} 0.03^{1} + \binom{18}{18} 0.97^{18} 0.03^{0} \\ &= 18 \times 0.595826 \times 0.03 + 1 \times 0.577951 \times 1 \\ &= 0.322 + 0.578 = 0.900 \end{aligned}$$

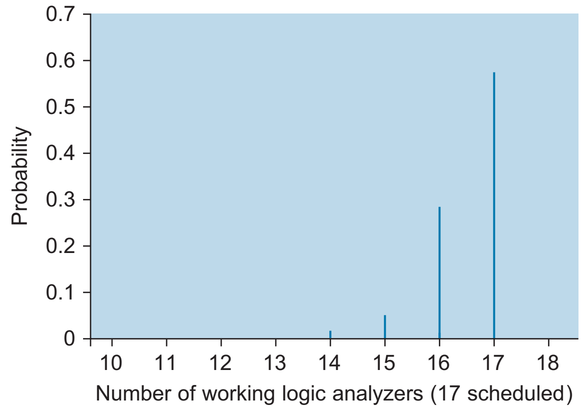


FIGURE 7.2.3 The probability distribution of the number of working logic analyzers produced if you plan to produce only 17. This is binomial, with $n = 17$ and $\pi = 0.97$.

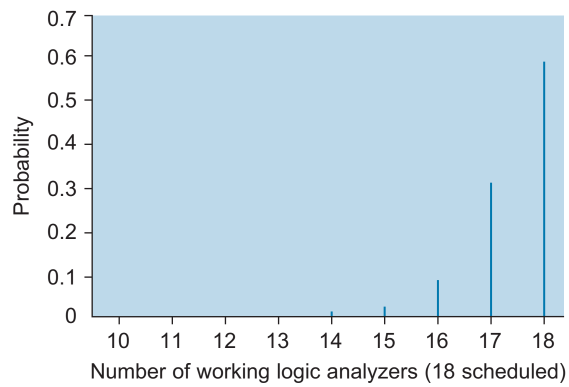


FIGURE 7.2.4 The probability distribution of the number of working logic analyzers produced *if you plan to produce 18*. This is binomial, with $n = 18$ and $\pi = 0.97$.

So if you schedule 18 for production, you have a 90% chance of shipping 17 good machines. It looks likely, but you would still be taking a 10% chance of failure. This probability distribution is shown in Figure 7.2.4.

Similar tedious calculations reveal that if you schedule 19 machines for production, you have a 98.2% chance of shipping 17 good machines (9.2% + 32.9% + 56.1%). So, to be reasonably sure of success, you'd better schedule at least 19 machines to get 17 good ones!

7.3 THE NORMAL DISTRIBUTION

You already know from Chapter 3 how to tell if a data set is approximately normally distributed. Now it's time to learn how to compute probabilities for this familiar bell-shaped distribution. One reason the normal distribution is particularly useful is the fact that, given only a mean and a standard deviation, you can compute any

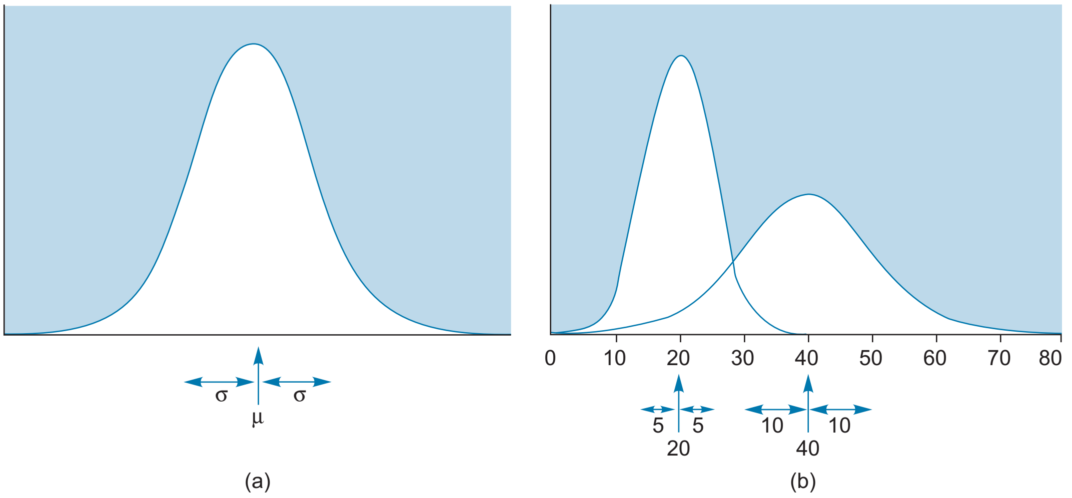


FIGURE 7.3.1 a. The normal distribution, with mean value μ and standard deviation σ. Note that the mean can be any number, and the standard deviation can be any positive number. b. Two different normal distributions. The one on the left has a smaller mean value (20) and a smaller standard deviation (5) than the other. The one on the right has mean 40 and standard deviation 10.

probability of interest (provided that the distribution really is normal).

The **normal distribution**, a continuous distribution, is represented by the familiar bell-shaped curve shown in Figure 7.3.1a. Note that there is a normal distribution for each combination of a mean value and a positive standard deviation value.[8] Just slide the curve to the right or left until the peak is centered above the mean value; then stretch it wider or narrower until the scale matches the standard deviation. Two different normal distributions are shown in Figure 7.3.1b.

Visualize Probabilities as the Area under the Curve

The bell-shaped curve gives you a guide for visualizing the probabilities for a normal distribution. You are more likely to see values occurring near the middle, where the curve is high. At the edges, where the curve is lower, values are not as likely to occur. Formally, it is the *area under the curve* that gives you the probability of being within a region, as illustrated in Figure 7.3.2.

Note that a shaded strip near the middle of the curve will have a larger area than a strip of the same width located nearer to the edge. Compare Figure 7.3.2 to Figure 7.3.3 to see this.

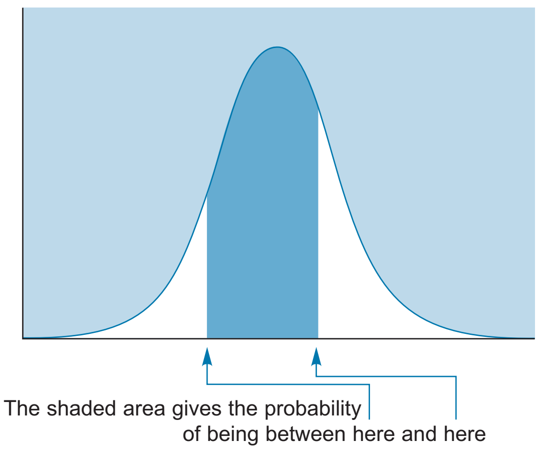


FIGURE 7.3.2 The probability that a normally distributed random variable is between any two values is equal to the area under the normal curve between these two values. You are more likely to see values in regions close to the mean.

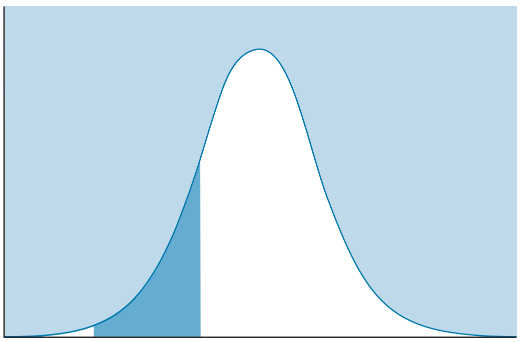


FIGURE 7.3.3 The probability of falling within a region that is farther from the middle of the curve. Since the normal curve is lower here, the probability is smaller than that shown in Figure 7.3.2.

8. The formula for the normal probability distribution with mean μ and standard deviation σ is

$$\frac{1}{\sqrt{2\pi}\sigma}e^{-[(x-\mu)/\sigma]^2/2}.$$

The Standard Normal Distribution Z and Its Probabilities

The **standard normal distribution** is a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. The letter Z is often used to denote a random variable that follows this standard normal distribution. One way to compute probabilities for a normal distribution is to use tables that give probabilities for the standard one, since it would be impossible to keep different tables for each combination of mean and standard deviation. The standard normal distribution can represent any normal distribution, provided you think in terms of the *number of standard deviations above or below the mean* instead of the actual units (e.g., dollars) of the situation. The standard normal distribution is shown in Figure 7.3.4.

The **standard normal probability table**, shown in Table 7.3.1, gives the probability that a standard normal random variable Z is *less than* any given number z. For example, the probability of being less than 1.38 is 0.9162, illustrated as an area in Figure 7.3.5. Doesn't it look like about 90% of the area? To find this number (0.9162), look up the value $z = 1.38$ in the standard normal probability table. While you're at it, look up 2.35 (to find 0.9906), 0 (to find 0.5000), and –0.82 (to find 0.2061). What is the probability corresponding to the value $z = 0.36$?

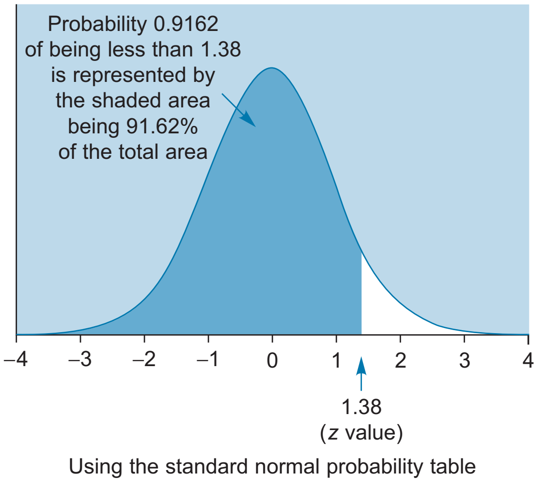


FIGURE 7.3.5 The probability that a standard normal random variable is *less than* $z = 1.38$ is 0.9162, as found in the standard normal probability table. This corresponds to the shaded region to the left of 1.38, which is 91.62% of the total area under the curve.

Solving Word Problems for Normal Probabilities

A typical word problem involving a normal distribution is a story involving some application to business that gives you a value for the mean and one for the standard deviation. Then you are asked to find one or more probabilities of interest. Here is an example of such a word problem:

The upper-management people at Simplified Technologies, Inc., have finally noticed that sales forecasts are usually wrong. Last quarter's sales were forecast as \$18 million but came in at \$21.3 million. Sales for the next quarter are forecast as \$20 million, with a standard deviation (based on previous experience) of \$3 million. Assuming a normal distribution centered at the forecast value, find the probability of a "really bad quarter," which is defined as sales lower than \$15 million.

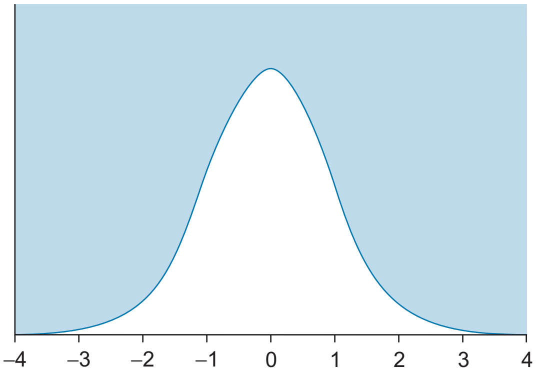


FIGURE 7.3.4 The standard normal distribution Z with mean value $\mu = 0$ and standard deviation $\sigma = 1$. The standard normal distribution may be used to represent any normal distribution, provided you think in terms of the number of standard deviations above or below the mean.

The beginning part sets the stage. The first numbers (18 and 21.3) describe past events but play no further role here. Instead, you should focus on the following facts:

- There is a normal distribution involved here.
- Its mean is μ = \$20 million.
- Its standard deviation is σ = \$3 million.
- You are asked to find the probability that sales will be lower than \$15 million.

The next step is to convert all of these numbers (except for the mean and standard deviation) into standardized numbers; this has to be done before you can look up the answer in the standard normal probability table. A **standardized number** (often written as z) is the number of standard deviations above the mean (or below the mean, if the standardized number is negative). This conversion is done as follows:

$$z = \text{Standardized number} = \frac{\text{Number} - \text{Mean}}{\text{Standard deviation}} = \frac{\text{Number} - \mu}{\sigma}$$

In this example, the number \$15 million is standardized as follows:

$$z = \frac{15 - \mu}{\sigma} = \frac{15 - 20}{3} = -1.67$$

TABLE 7.3.1 Standard Normal Probability Table (See Figure 7.3.5)

z Value	Probability	z Value	Probability	z Value	Probability	z Value	Probability	z Value	Probability	z Value	Probability
–2.00	0.0228	–1.00	0.1587	0.00	0.5000	0.00	0.5000	1.00	0.8413	2.00	0.9772
–2.01	0.0222	–1.01	0.1562	–0.01	0.4960	0.01	0.5040	1.01	0.8438	2.01	0.9778
–2.02	0.0217	–1.02	0.1539	–0.02	0.4920	0.02	0.5080	1.02	0.8461	2.02	0.9783
–2.03	0.0212	–1.03	0.1515	–0.03	0.4880	0.03	0.5120	1.03	0.8485	2.03	0.9788
–2.04	0.0207	–1.04	0.1492	–0.04	0.4840	0.04	0.5160	1.04	0.8508	2.04	0.9793
–2.05	0.0202	–1.05	0.1469	–0.05	0.4801	0.05	0.5199	1.05	0.8531	2.05	0.9798
–2.06	0.0197	–1.06	0.1446	–0.06	0.4761	0.06	0.5239	1.06	0.8554	2.06	0.9803
–2.07	0.0192	–1.07	0.1423	–0.07	0.4721	0.07	0.5279	1.07	0.8577	2.07	0.9808
–2.08	0.0188	–1.08	0.1401	–0.08	0.4681	0.08	0.5319	1.08	0.8599	2.08	0.9812
–2.09	0.0183	–1.09	0.1379	–0.09	0.4641	0.09	0.5359	1.09	0.8621	2.09	0.9817
–2.10	0.0179	–1.10	0.1357	–0.10	0.4602	0.10	0.5398	1.10	0.8643	2.10	0.9821
–2.11	0.0174	–1.11	0.1335	–0.11	0.4562	0.11	0.5438	1.11	0.8665	2.11	0.9826
–2.12	0.0170	–1.12	0.1314	–0.12	0.4522	0.12	0.5478	1.12	0.8686	2.12	0.9830
–2.13	0.0166	–1.13	0.1292	–0.13	0.4483	0.13	0.5517	1.13	0.8708	2.13	0.9834
–2.14	0.0162	–1.14	0.1271	–0.14	0.4443	0.14	0.5557	1.14	0.8729	2.14	0.9838
–2.15	0.0158	–1.15	0.1251	–0.15	0.4404	0.15	0.5596	1.15	0.8749	2.15	0.9842
–2.16	0.0154	–1.16	0.1230	–0.16	0.4364	0.16	0.5636	1.16	0.8770	2.16	0.9846
–2.17	0.0150	–1.17	0.1210	–0.17	0.4325	0.17	0.5675	1.17	0.8790	2.17	0.9850
–2.18	0.0146	–1.18	0.1190	–0.18	0.4286	0.18	0.5714	1.18	0.8810	2.18	0.9854
–2.19	0.0143	–1.19	0.1170	–0.19	0.4247	0.19	0.5753	1.19	0.8830	2.19	0.9857
–2.20	0.0139	–1.20	0.1151	–0.20	0.4207	0.20	0.5793	1.20	0.8849	2.20	0.9861
–2.21	0.0136	–1.21	0.1131	–0.21	0.4168	0.21	0.5832	1.21	0.8869	2.21	0.9864
–2.22	0.0132	–1.22	0.1112	–0.22	0.4129	0.22	0.5871	1.22	0.8888	2.22	0.9868
–2.23	0.0129	–1.23	0.1093	–0.23	0.4090	0.23	0.5910	1.23	0.8907	2.23	0.9871
–2.24	0.0125	–1.24	0.1075	–0.24	0.4052	0.24	0.5948	1.24	0.8925	2.24	0.9875
–2.25	0.0122	–1.25	0.1056	–0.25	0.4013	0.25	0.5987	1.25	0.8944	2.25	0.9878

TABLE 7.3.1 Standard Normal Probability Table (See Figure 7.3.5)—cont'd

z Value	Probability	*z* Value	Probability	*z* Value	Probability	*z* Value	Probability	*z* Value	Probability	*z* Value	Probability
–2.26	0.0119	–1.26	0.1038	–0.26	0.3974	0.26	0.6026	1.26	0.8962	2.26	0.9881
–2.27	0.0116	–1.27	0.1020	–0.27	0.3936	0.27	0.6064	1.27	0.8980	2.27	0.9884
–2.28	0.0113	–1.28	0.1003	–0.28	0.3897	0.28	0.6103	1.28	0.8997	2.28	0.9887
–2.29	0.0110	–1.29	0.0985	–0.29	0.3859	0.29	0.6141	1.29	0.9015	2.29	0.9890
–2.30	0.0107	–1.30	0.0968	–0.30	0.3821	0.30	0.6179	1.30	0.9032	2.30	0.9893
–2.31	0.0104	–1.31	0.0951	–0.31	0.3783	0.31	0.6217	1.31	0.9049	2.31	0.9896
–2.32	0.0102	–1.32	0.0934	–0.32	0.3745	0.32	0.6255	1.32	0.9066	2.32	0.9898
–2.33	0.0099	–1.33	0.0918	–0.33	0.3707	0.33	0.6293	1.33	0.9082	2.33	0.9901
–2.34	0.0096	–1.34	0.0901	–0.34	0.3669	0.34	0.6331	1.34	0.9099	2.34	0.9904
–2.35	0.0094	–1.35	0.0885	–0.35	0.3632	0.35	0.6368	1.35	0.9115	2.35	0.9906
–2.36	0.0091	–1.36	0.0869	–0.36	0.3594	0.36	0.6406	1.36	0.9131	2.36	0.9909
–2.37	0.0089	–1.37	0.0853	–0.37	0.3557	0.37	0.6443	1.37	0.9147	2.37	0.9911
–2.38	0.0087	–1.38	0.0838	–0.38	0.3520	0.38	0.6480	1.38	0.9162	2.38	0.9913
–2.39	0.0084	–1.39	0.0823	–0.39	0.3483	0.39	0.6517	1.39	0.9177	2.39	0.9916
–2.40	0.0082	–1.40	0.0808	–0.40	0.3446	0.40	0.6554	1.40	0.9192	2.40	0.9918
–2.41	0.0080	–1.41	0.0793	–0.41	0.3409	0.41	0.6591	1.41	0.9207	2.41	0.9920
–2.42	0.0078	–1.42	0.0778	–0.42	0.3372	0.42	0.6628	1.42	0.9222	2.42	0.9922
–2.43	0.0075	–1.43	0.0764	–0.43	0.3336	0.43	0.6664	1.43	0.9236	2.43	0.9925
–2.44	0.0073	–1.44	0.0749	–0.44	0.3300	0.44	0.6700	1.44	0.9251	2.44	0.9927
–2.45	0.0071	–1.45	0.0735	–0.45	0.3264	0.45	0.6736	1.45	0.9265	2.45	0.9929
–2.46	0.0069	–1.46	0.0721	–0.46	0.3228	0.46	0.6772	1.46	0.9279	2.46	0.9931
–2.47	0.0068	–1.47	0.0708	–0.47	0.3192	0.47	0.6808	1.47	0.9292	2.47	0.9932
–2.48	0.0066	–1.48	0.0694	–0.48	0.3156	0.48	0.6844	1.48	0.9306	2.48	0.9934
–2.49	0.0064	–1.49	0.0681	–0.49	0.3121	0.49	0.6879	1.49	0.9319	2.49	0.9936
–2.50	0.0062	–1.50	0.0668	–0.50	0.3085	0.50	0.6915	1.50	0.9332	2.50	0.9938
–2.51	0.0060	–1.51	0.0655	–0.51	0.3050	0.51	0.6950	1.51	0.9345	2.51	0.9940

TABLE 7.3.1 Standard Normal Probability Table (See Figure 7.3.5)—cont'd

z Value	Probability	*z* Value	Probability	*z* Value	Probability	*z* Value	Probability	*z* Value	Probability	*z* Value	Probability
–2.52	0.0059	–1.52	0.0643	–0.52	0.3015	0.52	0.6985	1.52	0.9357	2.52	0.9941
–2.53	0.0057	–1.53	0.0630	–0.53	0.2981	0.53	0.7019	1.53	0.9370	2.53	0.9943
–2.54	0.0055	–1.54	0.0618	–0.54	0.2946	0.54	0.7054	1.54	0.9382	2.54	0.9945
–2.55	0.0054	–1.55	0.0606	–0.55	0.2912	0.55	0.7088	1.55	0.9394	2.55	0.9946
–2.56	0.0052	–1.56	0.0594	–0.56	0.2877	0.56	0.7123	1.56	0.9406	2.56	0.9948
–2.57	0.0051	–1.57	0.0582	–0.57	0.2843	0.57	0.7157	1.57	0.9418	2.57	0.9949
–2.58	0.0049	–1.58	0.0571	–0.58	0.2810	0.58	0.7190	1.58	0.9429	2.58	0.9951
–2.59	0.0048	–1.59	0.0559	–0.59	0.2776	0.59	0.7224	1.59	0.9441	2.59	0.9952
–2.60	0.0047	–1.60	0.0548	–0.60	0.2743	0.60	0.7257	1.60	0.9452	2.60	0.9953
–2.61	0.0045	–1.61	0.0537	–0.61	0.2709	0.61	0.7291	1.61	0.9463	2.61	0.9955
–2.62	0.0044	–1.62	0.0526	–0.62	0.2676	0.62	0.7324	1.62	0.9474	2.62	0.9956
–2.63	0.0043	–1.63	0.0516	–0.63	0.2643	0.63	0.7357	1.63	0.9484	2.63	0.9957
–2.64	0.0041	–1.64	0.0505	–0.64	0.2611	0.64	0.7389	1.64	0.9495	2.64	0.9959
–2.65	0.0040	–1.65	0.0495	–0.65	0.2578	0.65	0.7422	1.65	0.9505	2.65	0.9960
–2.66	0.0039	–1.66	0.0485	–0.66	0.2546	0.66	0.7454	1.66	0.9515	2.66	0.9961
–2.67	0.0038	–1.67	0.0475	–0.67	0.2514	0.67	0.7486	1.67	0.9525	2.67	0.9962
–2.68	0.0037	–1.68	0.0465	–0.68	0.2483	0.68	0.7517	1.68	0.9535	2.68	0.9963
–2.69	0.0036	–1.69	0.0455	–0.69	0.2451	0.69	0.7549	1.69	0.9545	2.69	0.9964
–2.70	0.0035	–1.70	0.0446	–0.70	0.2420	0.70	0.7580	1.70	0.9554	2.70	0.9965
–2.71	0.0034	–1.71	0.0436	–0.71	0.2389	0.71	0.7611	1.71	0.9564	2.71	0.9966
–2.72	0.0033	–1.72	0.0427	–0.72	0.2358	0.72	0.7642	1.72	0.9573	2.72	0.9967
–2.73	0.0032	–1.73	0.0418	–0.73	0.2327	0.73	0.7673	1.73	0.9582	2.73	0.9968
–2.74	0.0031	–1.74	0.0409	–0.74	0.2296	0.74	0.7704	1.74	0.9591	2.74	0.9969
–2.75	0.0030	–1.75	0.0401	–0.75	0.2266	0.75	0.7734	1.75	0.9599	2.75	0.9970
–2.76	0.0029	–1.76	0.0392	–0.76	0.2236	0.76	0.7764	1.76	0.9608	2.76	0.9971
–2.77	0.0028	–1.77	0.0384	–0.77	0.2206	0.77	0.7794	1.77	0.9616	2.77	0.9972

TABLE 7.3.1 Standard Normal Probability Table (See Figure 7.3.5)—cont'd

z Value	Probability	*z* Value	Probability	*z* Value	Probability	*z* Value	Probability	*z* Value	Probability	*z* Value	Probability
–2.78	0.0027	–1.78	0.0375	–0.78	0.2177	0.78	0.7823	1.78	0.9625	2.78	0.9973
–2.79	0.0026	–1.79	0.0367	–0.79	0.2148	0.79	0.7852	1.79	0.9633	2.79	0.9974
–2.80	0.0026	–1.80	0.0359	–0.80	0.2119	0.80	0.7881	1.80	0.9641	2.80	0.9974
–2.81	0.0025	–1.81	0.0351	–0.81	0.2090	0.81	0.7910	1.81	0.9649	2.81	0.9975
–2.82	0.0024	–1.82	0.0344	–0.82	0.2061	0.82	0.7939	1.82	0.9656	2.82	0.9976
–2.83	0.0023	–1.83	0.0336	–0.83	0.2033	0.83	0.7967	1.83	0.9664	2.83	0.9977
–2.84	0.0023	–1.84	0.0329	–0.84	0.2005	0.84	0.7995	1.84	0.9671	2.84	0.9977
–2.85	0.0022	–1.85	0.0322	–0.85	0.1977	0.85	0.8023	1.85	0.9678	2.85	0.9978
–2.86	0.0021	–1.86	0.0314	–0.86	0.1949	0.86	0.8051	1.86	0.9686	2.86	0.9979
–2.87	0.0021	–1.87	0.0307	–0.87	0.1922	0.87	0.8078	1.87	0.9693	2.87	0.9979
–2.88	0.0020	–1.88	0.0301	–0.88	0.1894	0.88	0.8106	1.88	0.9699	2.88	0.9980
–2.89	0.0019	–1.89	0.0294	–0.89	0.1867	0.89	0.8133	1.89	0.9706	2.89	0.9981
–2.90	0.0019	–1.90	0.0287	–0.90	0.1841	0.90	0.8159	1.90	0.9713	2.90	0.9981
–2.91	0.0018	–1.91	0.0281	–0.91	0.1814	0.91	0.8186	1.91	0.9719	2.91	0.9982
–2.92	0.0018	–1.92	0.0274	–0.92	0.1788	0.92	0.8212	1.92	0.9726	2.92	0.9982
–2.93	0.0017	–1.93	0.0268	–0.93	0.1762	0.93	0.8238	1.93	0.9732	2.93	0.9983
–2.94	0.0016	–1.94	0.0262	–0.94	0.1736	0.94	0.8264	1.94	0.9738	2.94	0.9984
–2.95	0.0016	–1.95	0.0256	–0.95	0.1711	0.95	0.8289	1.95	0.9744	2.95	0.9984
–2.96	0.0015	–1.96	0.0250	–0.96	0.1685	0.96	0.8315	1.96	0.9750	2.96	0.9985
–2.97	0.0015	–1.97	0.0244	–0.97	0.1660	0.97	0.8340	1.97	0.9756	2.97	0.9985
–2.98	0.0014	–1.98	0.0239	–0.98	0.1635	0.98	0.8365	1.98	0.9761	2.98	0.9986
–2.99	0.0014	–1.99	0.0233	–0.99	0.1611	0.99	0.8389	1.99	0.9767	2.99	0.9986
–3.00	0.0013	–2.00	0.0228	–1.00	0.1587	1.00	0.8413	2.00	0.9772	3.00	0.9987

This (z = –1.67) tells you that \$15 million is 1.67 standard deviations below the mean (the forecast value).[9] Your problem has now been reduced to finding a standard normal probability:

Find the probability that a standard normal variable is less than $z = -1.67$.

From the table, you find the answer:

The probability of a really bad quarter is 0.0475, or about 5%.

Whew! It seems that a really bad quarter is not very likely. However, a 5% chance is an outside possibility that should not be disregarded altogether.

Figures 7.3.6 and 7.3.7 show this probability calculation, both in terms of sales dollars and in terms of standardized sales numbers (standard deviations above or below the mean).

This was an easy problem, since the answer was found directly from the standard normal probability table. Here is a question that requires a little more care:

Continuing with the sales-forecasting problem, find the probability of a "really good quarter," which is defined as sales in excess of \$24 million.

The first step is to standardize the sales number: \$24 million is $z = (24 - 20)/3 = 1.33$ standard deviations above the mean. Thus, you are asked to solve the following problem:

Find the probability that a standard normal variable *exceeds* $z = 1.33$.

Using the complement rule, we know that this probability is 1 minus the probability of being less than $z = 1.33$. Looking up 1.33 in the table, you find the answer:

$$\text{Probability of a really good quarter} = 1 - 0.9082$$
$$= 0.0918, \text{ or about } 9\%$$

This probability is illustrated, in standardized numbers, in Figure 7.3.8.

Here's another kind of problem:

Continuing with the sales-forecasting problem, find the probability of a "typical quarter," which is defined as sales between \$16 million and \$23 million.

Begin by standardizing both of these numbers, to see that your task is to solve the following problem:

Find the probability that a standard normal is *between* $z_1 = -1.33$ and $z_2 = 1.00$.

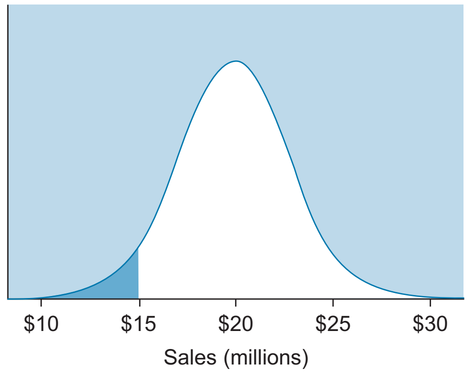


FIGURE 7.3.6 The probability of a really bad quarter (sales less than \$15 million) is represented by the shaded area under the curve. This is based on the forecast of \$20 million and the standard deviation of \$3 million. The answer is found by standardizing and then using the standard normal probability table.

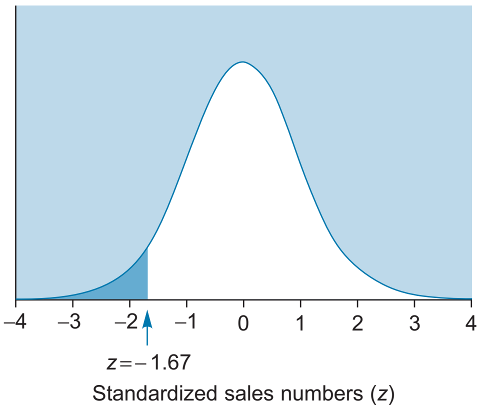


FIGURE 7.3.7 The probability of a really bad quarter, in terms of standardized sales numbers. This is the probability that sales will be more than $z = -1.67$ standard deviations below the mean. The answer is 0.0475.

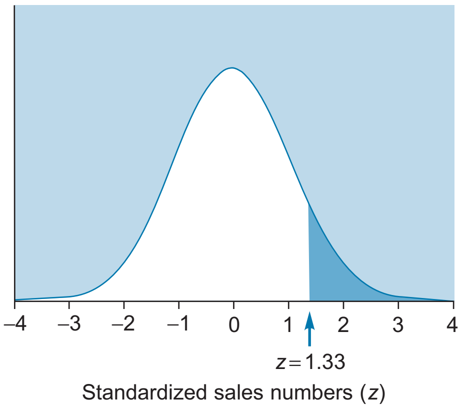


FIGURE 7.3.8 The probability of a really good quarter, in terms of standardized sales numbers. The shaded area is 1 minus the unshaded area under the curve, which may be looked up in the table. The answer is 0.0918.

9. You know that this is below the mean because the standardized number $z = -1.67$ is negative. The standardized number z will be positive for any number above the mean. The standardized number z for the mean itself is 0.

To solve this kind of problem, look up each standardized number in the table and find the difference between the probabilities for the answer. Be sure to subtract the smaller from the larger so that your answer is a positive number and therefore a "legal" probability!

$$\text{Probability of a typical quarter} = 0.8413 - 0.0918$$
$$= 0.7495, \text{ or about } 75\%$$

This probability is illustrated, in standardized numbers, in Figure 7.3.9.

Finally, here's yet another kind of problem:

> Continuing with the sales-forecasting problem, find the probability of a "surprising quarter," which is defined as sales either less than \$16 million or more than \$23 million.

This asks for the probability of *not* being between two numbers. Using the complement rule, you may simply take 1 minus the probability found in the preceding example, which was the probability of being between these two values. The answer is therefore as follows:

$$\text{Probability of a surprising quarter} = 1 - 0.7495$$
$$= 0.2505, \text{ or about } 25\%$$

This probability is illustrated, in standardized numbers, in Figure 7.3.10.

To use Excel® to compute these first three probabilities, we use the function "NORMDIST(value,mean,standardDeviation,TRUE)" to find the probability that a normal distribution with specified mean and standard deviation is less than some value. There is no need to standardize because Excel will do this for you as part of the calculation. The first calculation is straightforward because it is a probability of being less. The second calculation is one minus the NORMDIST function because it is a probability of being greater. The third calculation is the difference of two NORMDIST calculations because it is the probability of being between two values. Here are the results:

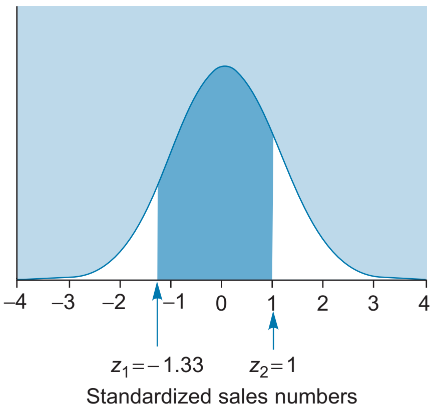


FIGURE 7.3.9 The probability of a typical quarter in terms of standardized sales numbers. The shaded area is found by looking up each standardized number in the table and then subtracting. Subtracting eliminates the *unshaded* area at the far left. The answer is 0.7495.

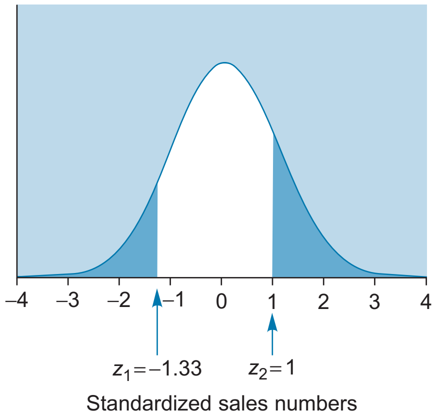
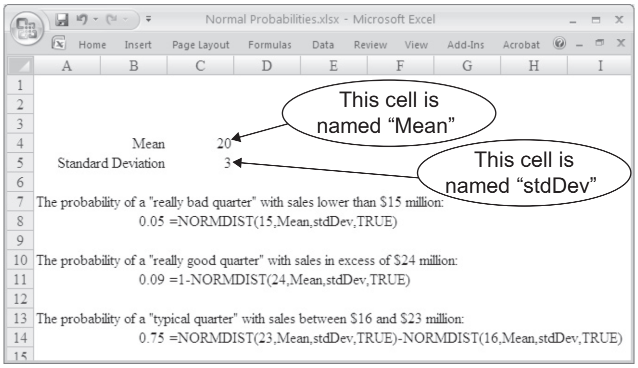


FIGURE 7.3.10 The probability of a surprising quarter, in terms of standardized sales numbers. The shaded area is found by looking up each standardized number in the table, finding the difference between the probabilities, and subtracting the result from 1. The answer is 0.2505.



The Four Different Probability Calculations

Here is a summary table of the four types of problems and how to solve them. The values z, z_1, and z_2 represent *standardized* numbers from the problem, found by subtracting the mean and dividing by the standard deviation. The table referred to is the standard normal probability table.

Computing Probabilities for a Normal Distribution

To Find the Probability of Being	Procedure
Less than z	Look up z in the table
More than z	Subtract above answer from 1
Between z_1 and z_2	Look up z_1 and z_2 in the table, and subtract smaller probability from larger
Not between z_1 and z_2	Subtract above answer (for "between z_1 and z_2") from 1

You may be wondering if there is a difference between the two events "sales exceeded \$22 million" and "sales were at least \$22 million." The term *exceeded* means *more than*, whereas the term *at least* means *more than or equal to*. In fact, for a normal distribution, there is *no difference* between the probabilities of these two events; the difference between the probabilities is just a geometric line, which represents no area under the normal curve.

Be Careful: Things Need Not Be Normal!

If you have a normal distribution and you know the mean and standard deviation, you can find correct probabilities by standardizing and then using the standard normal probability table. Fortunately, if the distribution is only approximately normal, your probabilities will still be approximately correct.

However, if the distribution is very far from normal, then any probabilities you might compute based on the mean, the standard deviation, and the normal table could be very wrong indeed.

Example

A Lottery (or Risky Project)

Consider a lottery (or a risky project, if you prefer) that pays back nothing 90% of the time, but pays \$500 the remaining 10% of the time. The expected (mean) payoff is \$50, and the standard deviation is \$150 for this discrete random variable. Note that this does *not* represent a normal distribution; it's not even close because it's so discrete, with only two possible values.

What is the probability of winning at least \$50? The correct answer is 10% because the *only* way to win anything at all is to win the full amount, \$500, which is at least \$50.

What if you assumed a normal distribution with this same mean (\$50) and standard deviation (\$150)? How far from the correct answer (10%) would you be? Very far away, because the probability that a normally distributed random variable exceeds its mean is 0.5 or 50%.

This is a big difference: 10% (the correct answer) versus 50% (computed by wrongly assuming a normal distribution). Figure 7.3.11 shows the large difference between the actual discrete distribution and the normal distribution with the same mean and standard deviation. Always be careful about assuming a normal distribution!

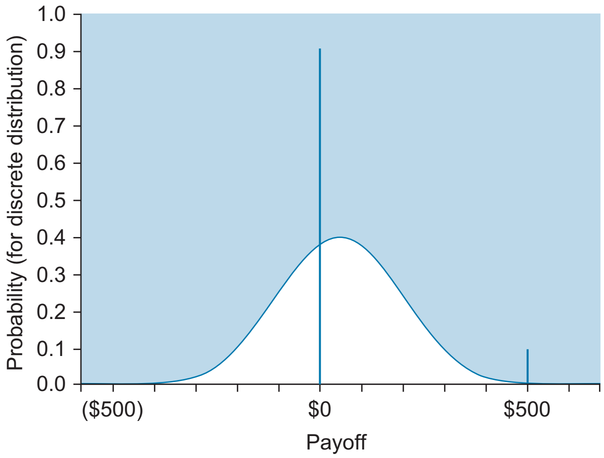


FIGURE 7.3.11 The discrete distribution of the payoff and the normal distribution having the same mean (\$50) and standard deviation (\$150). These distributions and their probabilities are very different. The discrete distribution gives the correct answers; the assumption of normality is wrong in this case.

7.4 THE NORMAL APPROXIMATION TO THE BINOMIAL

Remember the binomial distribution? It is the number of times that something happens out of n independent tries. A binomial distribution can never be exactly normal, for two reasons. First, any normal distribution is free to produce observations with decimal parts (e.g., 7.11327), whereas the binomial number X is restricted to whole numbers (e.g., 7). Second, a binomial distribution is skewed whenever π is any number other than 0.5 (becoming more and more skewed when π is close to 0 or to 1), whereas normal distributions are always perfectly symmetric.

However, a binomial distribution can be closely approximated by a normal distribution whenever the binomial n is large and the probability π is not too close to 0 or 1.[10,11] This fact will help you find probabilities (of being less, more, between, or not between given numbers) for a binomial distribution by replacing many complex and difficult calculations (using the earlier formula for individual binomial probabilities) with a single simpler calculation (using the normal distribution).

But how to choose a normal distribution that will be close to a given binomial distribution? A good choice is to use the normal distribution that has the same mean and standard deviation as the binomial distribution you wish to approximate. Since you already know how to find the mean and standard deviation for a binomial distribution (from Section 7.2), and you know how to find probabilities for a normal distribution with given mean

10. When π is close to 0 or 1, the approach to a normal distribution is slower as n increases due to skewness of the binomial with rare or nearly definite events. The Poisson distribution, covered in a later section, is a good approximation to the binomial when n is large and π is close to 0.

11. The central limit theorem, to be covered in Chapter 8, tells how a normal distribution emerges when many independent random trials are combined by adding or averaging.

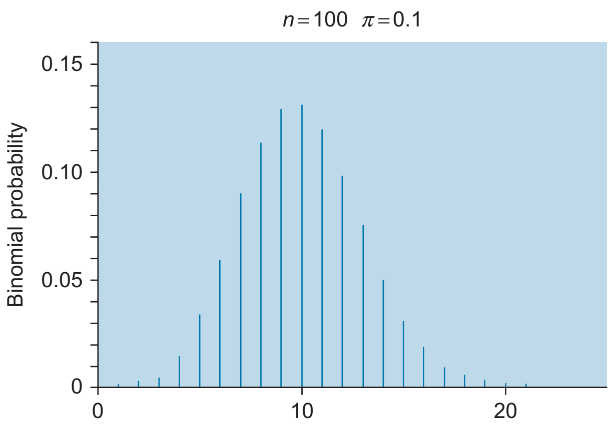


FIGURE 7.4.1 The probability distribution of a binomial with $n = 100$ and $\pi = 0.10$ is fairly close to normal.

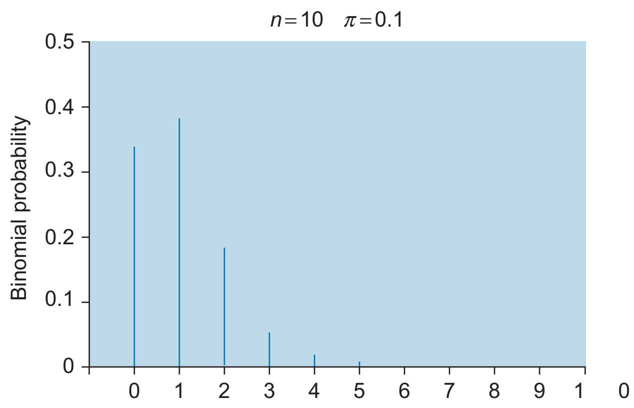


FIGURE 7.4.2 The probability distribution of a binomial with $n = 10$ and $\pi = 0.10$ is not very normal because n is not large enough.

and standard deviation (from Section 7.3), it should not be difficult for you to compute these approximate binomial probabilities.

Here is convincing evidence of the binomial approximation to the normal. Suppose n is 100 and π is 0.10. The probability distribution, computed using the binomial formula, is shown in Figure 7.4.1. It certainly has the bell shape of a normal distribution. Although it is still discrete, with separate, individual bars, there are enough observations that the discreteness is not a dominant feature.

To approximate a binomial (which is discrete, only taking on whole-number values) by using a normal random variable (which is continuous), we will do better if we extend the limits by one-half in each direction in order to include all numbers that round to the whole number(s).[12] For example, to approximate the probability that a binomial X is equal to 3, we would find the probability that a normal distribution (with the same mean and standard deviation) is between 2.5 and 3.5. We need to do this because the probability is zero that any normal random variable is equal to 3 and, actually, all values that the normal produced that were between 2.5 and 3.5 would round to 3. Similarly, to find the probability that a binomial is between 6 and 9, you would find the probability that a normal (with same mean and standard deviation) is between 5.5 and 9.5. The probability of *not* being between two numbers is, as usual, one minus the probability of being between them.

Using the Normal Approximation to the Binomial (Whole Numbers *a* and *b*):

The Probability That the Binomial Is:	Is Approximated by the Probability That the Corresponding Normal Is:
Exactly 8	Between 7.5 and 8.5
Exactly a	Between $a - 0.5$ and $a + 0.5$
Between 15 and 23	Between 14.5 and 23.5
Between a and b	Between $a - 0.5$ and $b + 0.5$

Compare Figure 7.4.1 ($n = 100$) to Figure 7.4.2 ($n = 10$) to see that, with smaller n, the distribution is not as normal. Furthermore, the discreteness is more important when n is small.

Example
High- and Low-Speed Microprocessors

We often don't have as much control over a manufacturing process as we would like. Such is the case with sophisticated microprocessor chips, such as some of those used in microcomputers, which can have over a billion transistors placed on a chip of silicon smaller than a square inch. Despite careful controls, there is variation within the resulting chips: Some will run at higher speeds than others.

In the spirit of the old software saying "It's not a bug, it's a feature!" the chips are sorted according to the speed at which they will actually run and priced accordingly (with the faster chips commanding a higher price). The catalog lists two products: 2 gigahertz (slower) and 3 gigahertz (faster).

Your machinery is known to produce the slow chips 80% of the time, on average, and fast chips the remaining 20% of the time, with chips being slow or fast independently of one another. Today your goal is to ship 1,000 slow chips and 300 fast chips, perhaps with some chips left over. How many should you schedule for production?

If you schedule 1,300 total chips, you expect 80% (1,040 chips) to be slow and 20% (260 chips) to be fast. You would

12. We assume here that you are looking for probabilities about a binomial *number of occurrences X*. If, on the other hand, you need probabilities for a binomial *proportion* or *percentage p*, you should first convert to X and then proceed from there. For example, the probability of observing "at least 20% of 261" is the same as the probability of observing "at least 53 of 261" since you need at least $0.20 \times 261 = 52.2$ and can observe only whole numbers.

have enough slow ones, but not enough fast ones, on average.

Since you know that you are limited by the number of fast chips, you compute 300/0.20 = 1,500. This tells you that if you schedule 1,500 chips, you can expect 20% of these (or 300 chips) to be fast. So on average, you would just meet the goal. Unfortunately, this means that you have only about a 50% chance of meeting the goal for fast chips!

Suppose you schedule 1,650 chips for production. What is the probability that you will be able to meet the goal? To solve this, you first state it as a complete probability question:

> Given a binomial random variable (the number of fast chips produced) with $n = 1{,}650$ total chips produced and $\pi = 0.20$ probability that a chip is fast, find the probability that this random variable is at least 300 but no more than 650.[13]

To solve this problem using the binomial distribution directly would require that you compute the probability for 300, for 301, for 302, and so on until you got tired. The normal approximation to the binomial allows you to solve it much faster using the standard normal probability table. You will need to know the mean and standard deviation of the number of fast chips produced:

$$\begin{aligned}\mu_{(\text{Number of fast chips})} &= n\pi \\ &= 1{,}650 \times 0.20 = 330\end{aligned}$$

$$\begin{aligned}\sigma_{(\text{Number of fast chips})} &= \sqrt{n\pi(1-\pi)} \\ &= \sqrt{1{,}650 \times 0.20 \times 0.80} = 16.24807\end{aligned}$$

You also need to standardize the bounds on the number of fast chips needed, 300 and 650 (after extending them by a half to 299.5 and 650.5), using the mean and standard deviation computed just above:

$$\begin{aligned}z_1 = \text{Standardized lower number of fast chips} &= \frac{299.5 - 330}{16.24807} \\ &= -1.88\end{aligned}$$

$$\begin{aligned}z_2 = \text{Standardized upper number of fast chips} &= \frac{650.5 - 330}{16.24807} \\ &= 19.73\end{aligned}$$

Looking up these standardized values in the standard normal probability table, you find 0.030 for $z_1 = -1.88$ and note that $z_2 = 19.73$ is way off the end of the table, so you use the value 1.[14] Subtracting these numbers to find the probability of being between the bounds, the answer is 1 – 0.030 = 0.970. You conclude that if 1,650 chips are scheduled, you have a 97% chance of meeting the goal of shipping 300 fast chips and 1,000 slow ones.

13. The reason is that more than 1,650 – 1,000 = 650 fast chips would imply fewer than 1,000 slow chips; thus, you would not be able to meet the goal for slow chips.

14. This makes sense; the probability that a standard normal variable will be less than $z_2 = 19.73$ standard deviations above its mean is essentially 1, since this nearly always happens.

One of the uses of probability is to help you understand what is happening "behind the scenes" in the real world. Let's see what might really be happening in an opinion poll by using a *What if* scenario analysis.

Example

Polling the Electorate

Your telephone polling and research firm was hired to conduct an opinion poll to see if a new municipal bond initiative is likely to be approved by the voters in the next election. You decided to interview 800 randomly selected representative people who are likely to vote, and you found that 437 intend to vote in favor. Here's the *What if*: If the entire electorate were, in fact, evenly divided on the issue, what is the probability that you would expect to see this many or more of your sample in favor?

> **Your coworker:** "It looks pretty close: 437 out of 800 is pretty close to 50–50, which would be 400 out of 800."
> **You:** "But 437 seems lots bigger than 400 to me. Let's find out if the extra 37 could reasonably be just randomness."
> **Your coworker:** "OK. Let's assume that each person is as likely to be in favor as not. Then we can compute the chances of seeing 437 or more."
> **You:** "OK. If the chances are more than 5% or 10%, then the extra 37 could reasonably be just randomness. But if the chances are really small, say under 5% or under 1%, then it would seem that more than just randomness is involved."

To do the calculation, let X represent the following binomial random variable: the number of people (out of 800 interviewed) who say they intend to vote in favor. If we assume that people are evenly divided on the issue, then the probability for each person interviewed is $\pi = 0.50$ that they intend to vote in favor. Now let's find the mean and standard deviation of X using formulas for the binomial distribution:

$$\mu_X = n\pi = (800)(0.50) = 400$$

$$\begin{aligned}\sigma_X &= \sqrt{n\pi(1-\pi)} \\ &= \sqrt{(800)(0.50)(1-0.50)} = 14.14214\end{aligned}$$

Now, to find the probability that X is at least 437, we extend the limit by one-half to find the probability that X is at least 436.5 and then use the fact that X is approximately normal. That is, we now find the probability that a normally distributed random variable with mean 400 and standard deviation 14.14214 is larger than 436.5 by standardizing as follows:

$$\begin{aligned}z = \text{Standardized value} &= \frac{436.5 - \mu_X}{\sigma_X} \\ &= \frac{436.5 - 400}{14.14214} = 2.58\end{aligned}$$

Using the normal probability tables, you find that the probability of seeing this large a margin (or larger) in favor

(*Continued*)

Example—cont'd

in the sample, assuming that the population is evenly divided, is 1 – 0.995 = 0.005. This is very unlikely: a probability of about a half a percent, or 1 out of 200.

You asked, what if the population were evenly divided, and found the answer: A sample percentage of 54.6% (this is 437/800) or more is highly unlikely. The conclusion is that you have evidence against the *What if* scenario of evenly divided voters. It looks good for the initiative!

7.5 TWO OTHER DISTRIBUTIONS: THE POISSON AND THE EXPONENTIAL

Many other probability distributions are useful in statistics. This section provides brief descriptions of two such distributions with an indication of how they might fit in with some general classes of business applications. The *Poisson distribution* is often useful as a model of the *number of events* that occur during a fixed time, such as arrivals. The *exponential distribution* can work well as a model of the *amount of time*, such as that required to complete an operation. These distributions work well together (as the *Poisson process*) with the Poisson representing the number of events and the exponential describing the time between events.

The Poisson Distribution

The Poisson distribution, like the binomial, is a counted number of times something happens. The difference is that there is no specified number n of possible tries. Here is one way that it can arise. If an event happens independently and randomly over time, and the mean rate of occurrence is constant over time, then the number of occurrences in a fixed amount of time will follow the **Poisson distribution**.[15] The Poisson is a *discrete* distribution and depends only on the mean number of occurrences expected.

Here are some random variables that might follow a Poisson distribution:

1. The number of orders your firm receives tomorrow.
2. The number of people who apply for a job tomorrow to your human resources division.
3. The number of defects in a finished product.
4. The number of calls your firm receives next week for help concerning an "easy-to-assemble" toy.
5. A binomial number X when n is large and π is small.

The following figures show what the Poisson probabilities look like for a system expecting a mean of 0.5 occurrence (Figure 7.5.1), 2 occurrences (Figure 7.5.2), and 20

15. *Poisson* is a French name, pronounced (more or less) "pwah-*soh*."

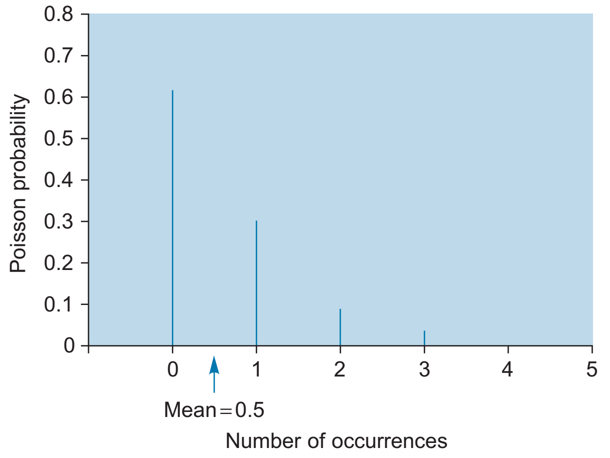


FIGURE 7.5.1 The Poisson distribution with 0.5 occurrences expected is a skewed distribution. There is a high probability, 0.607, that no occurrences will happen at all.

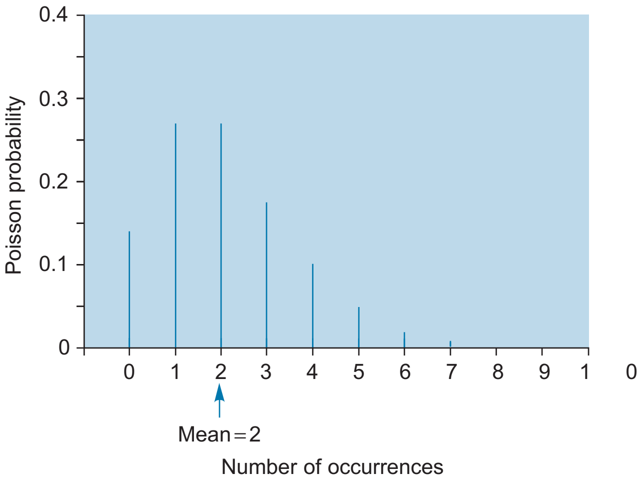


FIGURE 7.5.2 The Poisson distribution with two occurrences expected. The distribution is still somewhat skewed.

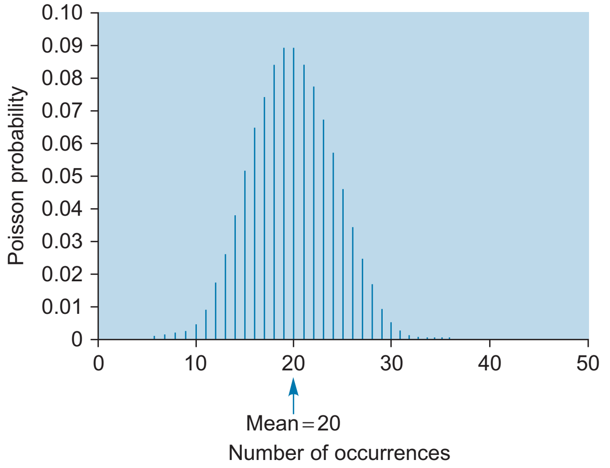


FIGURE 7.5.3 The Poisson distribution with 20 occurrences expected. The distribution, although still discrete, is now fairly close to normal.

occurrences (Figure 7.5.3). Note from the bell shape of Figure 7.5.3 that the Poisson distribution is approximately normal when many occurrences are expected.

There are three important facts about a Poisson distribution. These facts, taken together, tell you how to find probabilities for a Poisson distribution when you know only its mean.

For a Poisson Distribution

1. The standard deviation is always equal to the square root of the mean: $\sigma = \sqrt{\ }$.
2. The exact probability that a Poisson random variable X with mean is equal to a is given by the formula

$$P(X = a) = \frac{\mu^a}{a!} e^{-\mu}$$

where $e = 2.71828\ldots$ is a special number.[16]

3. If the mean is large, then the Poisson distribution is approximately normal.

16. This special mathematical number also shows up in continuously compounded interest formulas.

Example

How Many Warranty Returns?

Because your firm's quality is so high, you expect only 1.3 of your products to be returned, on average, each day for warranty repairs. What are the chances that no products will be returned tomorrow? That one will be returned? How about two? How about three?

Since the mean (1.3) is so small, exact calculations are needed. Here are the details:

$$P(X = 0) = \frac{1.3^0}{0!} e^{-1.3} = \frac{1}{1} \times 0.27253 = 0.27253$$

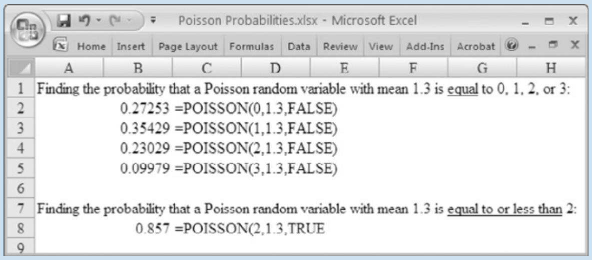
$$P(X = 1) = \frac{1.3^1}{1!} e^{-1.3} = \frac{1.3}{1} \times 0.27253 = 0.35429$$

$$P(X = 2) = \frac{1.3^2}{2!} e^{-1.3} = \frac{1.69}{2} \times 0.27253 = 0.23029$$

$$P(X = 3) = \frac{1.3^3}{3!} e^{-1.3} = \frac{2.197}{6} \times 0.27253 = 0.09979$$

From these basic probabilities, you could add up the appropriate probabilities for 0, 1, and 2 to also find the probability that two items *or fewer* will be returned. The probability is, then, $0.27253 + 0.35429 + 0.23029 = 0.857$, or 85.7%.

To use Excel to compute these probabilities, you could use the function "POISSON(value,mean,FALSE)" to find the probability that a Poisson random variable is exactly equal to some value, and you could use "POISSON(value, mean,TRUE)" to find the probability that a Poisson random variable is *less than or equal* to the value. Here are the results:



1	Finding the probability that a Poisson random variable with mean 1.3 is equal to 0, 1, 2, or 3:
2	0.27253 =POISSON(0,1.3,FALSE)
3	0.35429 =POISSON(1,1.3,FALSE)
4	0.23029 =POISSON(2,1.3,FALSE)
5	0.09979 =POISSON(3,1.3,FALSE)
6	
7	Finding the probability that a Poisson random variable with mean 1.3 is equal to or less than 2:
8	0.857 =POISSON(2,1.3,TRUE)
9	

Example

How Many Phone Calls?

Your firm handles 460 calls per day, on average. Assuming a Poisson distribution, find the probability that you will be overloaded tomorrow, with 500 or more calls received.

The mean, $\mu = 460$, is given. The standard deviation is $\sigma = 21.44761$. You may use the normal approximation because the mean (460) is so large. Since the normal distribution is continuous, any value over 499.5 will round to 500 or more. The standardized number of calls is

$$z = \frac{499.5 - \mu}{\sigma} = \frac{499.5 - 460}{21.44761} = 1.84$$

When you use the standard normal probability table, the answer is a probability of $1 - 0.967 = 0.033$, so you may expect to be overloaded tomorrow with probability only about 3% (not very likely but within possibility).

The Exponential Distribution

The **exponential distribution** is the very skewed continuous distribution shown in Figure 7.5.4. Its rise is vertical at 0, on the left, and it descends gradually, with a long tail on the right.

The following is a situation in which the exponential distribution is appropriate. If events happen independently

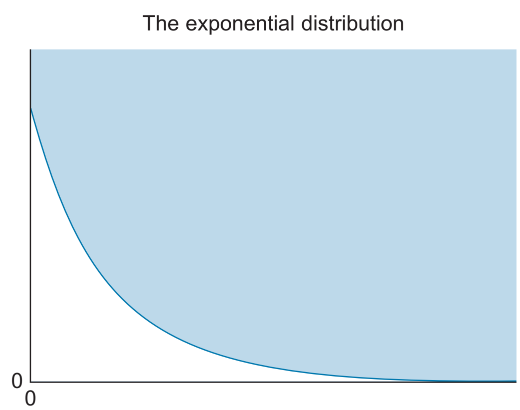


FIGURE 7.5.4 The exponential distribution is a very skewed distribution that is often used to represent waiting times between events.

and randomly with a constant rate over time, the *waiting time* between successive events follows an exponential distribution.[17]

Here are some examples of random variables that might follow an exponential distribution:

1. Time between customer arrivals at an auto repair shop.
2. The amount of time your copy machine works between visits by the repair people.
3. The length of time of a typical telephone call.
4. The time until a TV system fails.
5. The time it takes to provide service for one customer.

The exponential distribution *has no memory* in the surprising sense that after you have waited awhile without success for the next event, your mean waiting time remaining until the next event is no shorter than it was when you started! This makes sense for waiting times, since occurrences are independent of one another and "don't know" that none have happened recently.

What does this property say about telephone calls? Suppose you are responsible for a switching unit for which the average call lasts five minutes. Consider all calls received at a given moment. On average, you expect them to last five minutes, with the individual durations following the exponential distribution. After one minute passes, some of these calls have ended. However, the calls that remain are all expected, on average, to last *five minutes more*. The reason is that the shorter calls have already been eliminated. While this may be difficult to believe, it has been confirmed (approximately) using real data.

Here are the basic facts for an exponential distribution. Note that there is no "normal approximation" because the exponential distribution is *always* very skewed.

For an Exponential Distribution

1. The standard deviation is always equal to the mean: $\sigma = \mu$.
2. The exact probability that an exponential random variable X with mean μ is less than a is given by the formula

$$P(X \le a) = 1 - e^{-a/\mu}$$

There is a relationship between the exponential and the Poisson distributions when events happen independently at a constant rate over time. The *number of events* in any fixed time period is Poisson, and the *waiting time between events* is exponential. This is illustrated in Figure 7.5.5. In fact, the distribution of the waiting time from any fixed time until the next event is exponential.

17. Note that this implies that the total number of events follows a Poisson distribution.

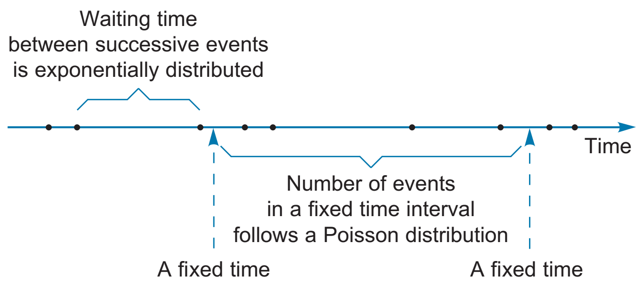


FIGURE 7.5.5 The relationship between the exponential and the Poisson distributions when events happen over time independently and at a constant rate.

Example
Customer Arrivals

Suppose customers arrive independently at a constant mean rate of 40 per hour. To find the probability that at least one customer arrives in the next five minutes, note that this is the probability that the exponential waiting time until the next customer arrives is less than five minutes. Since 40 customers arrive each hour, on average, the mean of this exponential random variable is $\mu = 1/40 = 0.025$ hours, or $0.025 \times 60 = 1.5$ minutes. The probability is then $P(X \le 5) = 1 - e^{-5/1.5} = 0.964$, which may be computed in Excel® using the formula "=1–EXP(–5/1.5)". So the chances are high (96.4%) that at least one customer will arrive in the next 5 minutes.

7.6 END-OF-CHAPTER MATERIALS

Summary

A **random variable** is a specification or description of a numerical result from a random experiment. A particular value taken on by a random variable is called an **observation**. The pattern of probabilities for a random variable is called its **probability distribution**. Random variables are either **discrete** (if you can list all possible outcomes) or **continuous** (if any number in a range is possible). Some random variables are actually discrete, but you can work with them as though they were continuous.

For a discrete random variable, the probability distribution is a list of the possible values together with their probabilities of occurrence. The mean or expected value and the standard deviation are computed as follows.

For a discrete random variable:

$$\mu = E(X) = \text{Sum of (value times probability)} = \sum XP(X)$$

$$\sigma = \sqrt{\text{Sum of (squared deviation times probability)}}$$

$$= \sqrt{\sum (X-\mu)^2 P(X)}$$

The interpretations are familiar. The **mean** or **expected value** indicates the typical or average value, and the

standard deviation indicates the risk in terms of approximately how far from the mean you can expect to be.

A random variable X has a **binomial distribution** if it represents the *number of occurrences* of an event out of n trials, provided (1) for each of the n trials, the event always has the same probability π of happening, and (2) the trials are independent of one another. The **binomial proportion** is $p = X/n$, which also represents a percentage. The mean and standard deviation of a binomial or binomial proportion may be found as follows:

Mean and Standard Deviation for a Binomial Distribution

	Number of Occurrences, X	Proportion or Percentage, $p = X/n$
Mean	$E(X) = \mu_X = n\pi$	$E(p) = \mu_p = \pi$
Standard deviation	$\sigma_X = \sqrt{n\pi(1-\pi)}$	$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$

The probability that a binomial random variable X is equal to some given number a (from 0 to n) is given by the following formula. Binomial probability that X equals a:

$$P(X=a) = \binom{n}{a}\pi^a(1-\pi)^{n-a}$$
$$= \frac{n!}{a!(n-a)!}\pi^a(1-\pi)^{n-a}$$
$$= \frac{1\times2\times3\times\cdots\times n}{(1\times2\times3\times\cdots\times a)[1\times2\times3\times\cdots\times(n-a)]}\pi^a(1-\pi)^{n-a}$$

The notation $n!$ is n factorial, the product of the numbers from 1 to n, with $0! = 1$ by definition. The notation

$$\binom{n}{a} = \frac{n!}{a!(n-a)!}$$

is the *binomial coefficient*, read aloud as "n choose a."

The **normal distribution**, a continuous distribution, is represented by the familiar bell-shaped curve. The probability that a normal random variable will be between any two values is equal to the area under the normal curve between these two values. There is a normal distribution for each combination of a mean μ and a (positive) standard deviation σ. The **standard normal distribution** is a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. You may think of the standard normal distribution as representing the number of standard deviations above or below the mean. The **standard normal probability table** gives the probability that a standard normal random variable Z is *less than* any given number z.

To solve word problems involving normal probabilities, first identify the mean μ, standard deviation σ, and the probability asked for. Convert to a **standardized number** z (the number of standard deviations above the mean, or below the mean if the standardized number is negative) by subtracting the mean and dividing by the standard deviation:

$$z = \text{Standardized number} = \frac{\text{Number} - \text{Mean}}{\text{Standard deviation}}$$
$$= \frac{\text{Number} - \mu}{\sigma}$$

Finally, look up the standardized number or numbers in the standard normal probability table and use the following summary table (where z, z_1, and z_2 are standardized numbers) to find the final answer:

Computing Probabilities for a Normal Distribution

To Find the Probability of Being	Procedure
Less than z	Look up z in the table
More than z	Subtract above answer from 1
Between z_1 and z_2	Look up z_1 and z_2 in the table, and subtract smaller probability from larger
Not between z_1 and z_2	Subtract above answer (for "between z_1 and z_2") from 1

Probabilities for a binomial distribution may be approximated using the normal distribution with the same mean and standard deviation, provided n is large and π is not too close to 0 or 1. Since the normal distribution is continuous, extend the limits by one-half in each direction (for example, the probability that a binomial is exactly a whole number a is approximated by the probability that the corresponding normal is between $a - 0.5$ and $a + 0.5$).

If occurrences happen independently and randomly over time, and the average rate of occurrence is constant over time, then the number of occurrences that happen in a fixed amount of time will follow the **Poisson distribution**, a discrete random variable. The standard deviation is the square root of the mean. If the mean is large, the Poisson distribution is approximately normal and the standard normal probability table may be used. Exact Poisson probabilities may be found using the following formula:

$$P(X = a) = \frac{\mu^a}{a!}e^{-\mu}$$

The **exponential distribution** is a very skewed continuous distribution useful for understanding such variables as waiting times and durations of telephone calls. It has no "memory," in the sense that after you have waited awhile without success for the next event, your average waiting time until the next event is no shorter than it was when you started. Its standard deviation is always equal to its mean. The probability that an exponential random variable X with mean μ is less than or equal to a is $P(X \le a) = 1 - e^{-a/\mu}$. There is no normal approximation for an exponential random variable.

Key Words

Questions

1. a. What is a random variable?
   b. What is the difference between a random variable and a number?
2. a. What is a discrete random variable?
   b. What is a continuous random variable?
   c. Give an example of a discrete random variable that is continuous for practical purposes.
3. a. What is the probability distribution of a discrete random variable?
   b. How do you find the mean of a discrete random variable? How do you interpret the result?
   c. How do you find the standard deviation of a discrete random variable? How do you interpret the result?
4. a. How do you tell if a random variable has a binomial distribution?
   b. What is a binomial proportion?
   c. What are n, π, X, and p?
5. For a binomial distribution:
   a. Why don't you just construct the probability tree to find the probabilities?
   b. How do you find the mean and the standard deviation?
   c. How do you find the probability that X is equal to some number?
   d. How do you find the exact probability that X is greater than or equal to some number?
   e. If n is a large number, how do you find the approximate probability that X is greater than or equal to some number?
6. a. What is a factorial?
   b. Find 3!, 0!, and 15!.
   c. What is a binomial coefficient? What does it represent in the formula for a binomial probability?
   d. Find the binomial coefficient "8 choose 5."
7. a. What is a normal distribution?
   b. Identify all of the different possible normal distributions.
   c. What does the area under the normal curve represent?
   d. What is the standard normal distribution? What is it used for?
   e. What numbers are found in the standard normal probability table?
   f. Find the probability that a standard normal random variable is less than –1.65.
   g. How do you standardize a number?
8. a. What kinds of situations give rise to a Poisson distribution?
   b. Is the Poisson a discrete or a continuous distribution?
   c. What is the standard deviation of a Poisson distribution?
   d. How do you find probabilities for a Poisson distribution if the mean is large?
   e. How do you find exact probabilities for a Poisson distribution?
9. a. What kinds of situations give rise to an exponential distribution?
   b. What is meant by the fact that an exponential random variable has no memory?
   c. Can the standard normal probability table be used to find probabilities for an exponential distribution? Why or why not?
   d. How do you find probabilities for an exponential distribution?

Problems

Problems marked with an asterisk (*) are solved in the Self Test in Appendix C

1. A call option on common stock is being evaluated. If the stock goes down, the option will expire worthless. If the stock goes up, the payoff depends on just how high the stock goes. For simplicity, the payoffs are modeled as a discrete distribution with the probability distribution in Table 7.6.1. Even though options markets, in fact, behave more like a continuous random variable, this discrete approximation will give useful approximate results. Answer the following questions based on the discrete probability distribution given.
   a.* Find the mean, or expected value, of the option payoff.
   b.* Describe briefly what this expected value represents.
   c.* Find the standard deviation of the option payoff.
   d.* Describe briefly what this standard deviation represents.
   e.* Find the probability that the option will pay at least $20.
   f. Find the probability that the option will pay less than $30.

TABLE 7.6.1 Probability Distribution of Payoff

Payoff	Probability
$0	0.50
10	0.25
20	0.15
30	0.10

TABLE 7.6.2 Probability Distribution of Downtime

Problem	Downtime (minutes)	Probability
Minor	5	0.60
Substantial	30	0.30
Catastrophic	120	0.10

2. The length of time a system is "down" (that is, broken) is described (approximately) by the probability distribution in Table 7.6.2. Assume that these downtimes are exact. That is, there are three types of easily recognized problems that always take this long (5, 30, or 120 minutes) to fix.
   a. What kind of probability distribution does this table represent?
   b. Find the mean downtime.
   c. Find the standard deviation of the downtime.
   d. What is the probability that the downtime will be greater than 10 minutes, according to this table?
   e. What is the probability that the downtime is literally within one standard deviation of its mean? Is this about what you would expect for a normal distribution?
3. An investment will pay $105 with probability 0.7, and $125 with probability 0.3. Find the risk (as measured by standard deviation) for this investment.
4. On a given day, assume that there is a 30% chance you will receive no orders, a 50% chance you will receive one order, a 15% chance of two orders, and a 5% chance of three orders. Find the expected number of orders and the variability in the number of orders.
5. A new project has an uncertain cash flow. A group meeting has resulted in a consensus that a reasonable way to view the possible risks and rewards is to say that the project will pay $50,000 with probability 0.2, will pay $100,000 with probability 0.3, will pay $200,000 with probability 0.4, and will pay $400,000 with probability 0.1. How much risk is involved here? Please give both the name and the numerical value of your answer.
6. Your company is hoping to fill a key technical position and has advertised in hopes of obtaining qualified applicants. Because of the demanding qualifications, the pool of qualified people is limited and Table 7.6.3 shows your subjective probabilities for each outcome.
   a. Find the probability of obtaining at least one applicant.
   b. Find the probability of obtaining two or more applicants.
   c. Find the mean number of applicants.
   d. Find the standard deviation of the number of applicants and write a sentence interpreting its meaning.
7. You work for the loan department of a large bank. You know that one of your customers has been having trouble with the recession and may not be able to make the loan payment that is due next week. You believe there is a 60% chance that the payment of $50,000 will be made in full, a 30% chance that only half will be paid, and a 10% chance that no payment will be made at all.
   a. Find the expected loan payment.
   b. Find the degree of risk for this situation.
8. You are planning to invest in a new high-tech company, and figure your rate of return over the coming year as in Table 7.6.4 (where 100% says that you doubled your money, –50% says you lost half, etc.).
   a. Find the mean rate of return and explain what it represents.
   b. Find the standard deviation of the rate of return and explain what it represents.
   c. Find the probability that you will earn more than 40%, according to the table.
   d. How would you measure the risk of this investment?

TABLE 7.6.3 Probabilities for Qualified Technical Applicants

Number of Applicants	Probability
0	0.30
1	0.55
2	0.10
3	0.05

TABLE 7.6.4 Rates of Return and Probabilities for Four Scenarios

Rate of Return	Probability
100%	0.20
50	0.40
0	0.25
–50	0.15

9. You can invest in just one of four projects on a lot of land you own. For simplicity, you have modeled the payoffs (as net present value in today's dollars) of the projects as discrete distributions. By selling the land, you can make \$60,000 for sure. If you build an apartment, you estimate a payoff of \$130,000 if things go well (with probability 0.60) and \$70,000 otherwise. If you build a single-family house, the payoff is \$100,000 (with probability 0.60) and \$60,000 otherwise. Finally, you could build a gambling casino which would pay very well—\$500,000—but with a probability of just 0.10 since the final government permits are not likely to be granted; all will be lost otherwise.

- **a.** Find the expected payoff for each of these four projects. In terms of just the expected payoff, rank these projects in order from best to worst.
- **b.** Find the standard deviation for each of these four projects. In terms of risk only, rank the projects from best to worst.
- **c.** Considering both the expected payoff and the risk involved, can any project or projects be eliminated from consideration entirely?
- **d.** How would you decide among the remaining projects? In particular, does any single project dominate the others completely?

10. Your quality control manager has identified the four major problems, the extent to which each one occurs (i.e., the probability that this problem occurs per item produced), and the cost of reworking to fix each one (see Table 7.6.5). Assume that only one problem can occur at a time.

- **a.** Compute the expected rework cost for each problem separately. For example, the expected rework cost for "broken case" is 0.04×6.88. Compare the results and indicate the most serious problem in terms of expected dollar costs.
- **b.** Find the overall expected rework cost due to all four problems together.
- **c.** Find the standard deviation of rework cost (don't forget the nonreworked items).
- **d.** Write a brief memo, as if to your supervisor, describing and analyzing the situation.

TABLE 7.6.5 **Quality Control Problems: Type, Extent, and Cost**

Problem	Probability	Rework Cost
Broken case	0.04	\$6.88
Faulty electronics	0.02	12.30
Missing connector	0.06	0.75
Blemish	0.01	2.92

11. Suppose that 8% of the loans you authorize as vice president of the consumer loan division of a neighborhood bank will never be repaid. Assume further that you authorized 284 loans last year and that loans go sour independently of one another.

- **a.** How many of these loans, authorized by you, do you expect will never be repaid? What percentage do you expect?
- **b.** Find the usual measure of the level of uncertainty in the number of loans you authorized that will never be repaid. Briefly interpret this number.
- **c.** Find the usual measure of the level of uncertainty in the percentage of loans you authorized that will never be repaid. Briefly interpret this number.

12. Your company is planning to market a new reading lamp and has segmented the market into three groups—avid readers, regular readers, and occasional readers—and currently assumes that 25% of avid readers, 15% of regular readers, and 10% of occasional readers will want to buy the new product. As part of a marketing survey, 400 individuals will be randomly selected from the population of regular readers. Using the current assumptions, find the mean and standard deviation of the percentage among those surveyed who will want to buy the new product.

13. A company is conducting a survey of 235 people to measure the level of interest in a new product. Assume that the probability of a randomly selected person's being "very interested" is 0.88 and that people are selected independently of one another.

- **a.** Find the standard deviation of the percentage who will be found by the survey to be very interested.
- **b.** How much uncertainty is there in the number of people who will be found to be very interested?
- **c.** Find the expected number of people in the sample who will say that they are very interested.
- **d.** Find the expected percentage that the survey will identify as being very interested.

14. An election coming up next week promises to be very close. In fact, assume that 50% are in favor and 50% are against. Suppose you conduct a poll of 791 randomly selected likely voters. Approximately how different will the percent in favor (from the poll) be from the 50% in the population you are trying to estimate?

15. Repeat the previous problem, but now assume that 85% are in favor in the population. Is the uncertainty larger or smaller than when 50% was assumed? Why?

16. You have just performed a survey interviewing 358 randomly selected people. You found that 94 of them are interested in possibly purchasing a new cable TV service. How much uncertainty is there in this number "94" as compared to the average number you would expect to find in such a survey? (You may assume that exactly 25% of all people you might have interviewed would have been interested.)

17. You are planning to make sales calls at eight firms today. As a rough approximation, you figure that each call has a 20% chance of resulting in a sale and that firms make their buying decisions without consulting each other. Find the probability of having a really terrible day with no sales at all.

18. It's been a bad day for the market, with 80% of securities losing value. You are evaluating a portfolio of 15 securities and will assume a binomial distribution for the number of securities that lost value.
   a.* What assumptions are being made when you use a binomial distribution in this way?
   b.* How many securities in your portfolio would you expect to lose value?
   c.* What is the standard deviation of the number of securities in your portfolio that lose value?
   d.* Find the probability that all 15 securities lose value.
   e.* Find the probability that exactly 10 securities lose value.
   f. Find the probability that 13 or more securities lose value.

19. Your firm has decided to interview a random sample of 10 customers in order to determine whether or not to change a consumer product. Your main competitor has already done a similar but much larger study and has concluded that exactly 86% of consumers approve of the change. Unfortunately, your firm does not have access to this information (but you may use this figure in your computations here).
   a. What is the name of the probability distribution of the number of consumers who will approve of the change in your study?
   b. What is the expected number of people, out of the 10 you will interview, who will approve of the change?
   c. What is the standard deviation of the number of people, out of the 10 you will interview, who will approve of the change?
   d. What is the expected percentage of people, out of the 10 you will interview, who will approve of the change?
   e. What is the standard deviation of the percentage of people, out of the 10 you will interview, who will approve of the change?
   f. What is the probability that exactly eight of your interviewed customers will approve of the change?
   g. What is the probability that eight or more of your interviewed customers will approve of the change?

20. Suppose that the number of hits on your company's website, from noon to 1 p.m. on a typical weekday, follows a normal distribution (approximately) with a mean of 190 and a standard deviation of 24.
   a. Find the probability that the number of hits is more than 160.
   b. Find the probability that the number of hits is less than 215.
   c. Find the probability that the number of hits is between 165 and 195.
   d. Find the probability that the number of hits is not between 150 and 225.

21. Find the probability that you will see moderate improvement in productivity, meaning an increase in productivity between 6 and 13. You may assume that the productivity increase follows a normal distribution with a mean of 10 and a standard deviation of 7.

22. Under usual conditions, a distillation unit in a refinery can process a mean of 135,000 barrels per day of crude petroleum, with a standard deviation of 6,000 barrels per day. You may assume a normal distribution.
   a. Find the probability that more than 135,000 barrels will be produced on a given day.
   b. Find the probability that more than 130,000 barrels will be produced on a given day.
   c. Find the probability that more than 150,000 barrels will be produced on a given day.
   d. Find the probability that less than 125,000 barrels will be produced on a given day.
   e. Find the probability that less than 100,000 barrels will be produced on a given day.

23. The quality control section of a purchasing contract for valves specifies that the diameter must be between 2.53 and 2.57 centimeters. Assume that the production equipment is set so that the mean diameter is 2.56 centimeters and the standard deviation is 0.01 centimeter. What percent of valves produced, over the long run, will be within these specifications, assuming a normal distribution?

24. Assume that the stock market closed at 13,246 points today. Tomorrow you expect the market to rise a mean of 4 points, with a standard deviation of 115 points. Assume a normal distribution.
   a. Find the probability that the stock market goes down tomorrow.
   b. Find the probability that the market goes up more than 50 points tomorrow.
   c. Find the probability that the market goes up more than 100 points tomorrow.
   d. Find the probability that the market goes down more than 150 points tomorrow.
   e. Find the probability that the market changes by more than 200 points in either direction.

25. Based on recent experience, you expect this Saturday's total receipts to have a mean of \$2,353.25 and a standard deviation of \$291.63 and to be normally distributed.
   a. Find the probability of a typical Saturday, defined as total receipts between \$2,000 and \$2,500.
   b. Find the probability of a terrific Saturday, defined as total receipts over \$2,500.
   c. Find the probability of a mediocre Saturday, defined as total receipts less than \$2,000.

26. The amount of ore (in tons) in a segment of a mine is assumed to follow a normal distribution with mean 185 and standard deviation 40. Find the probability that the amount of ore is less than 175 tons.

27. You are a farmer about to harvest your crop. To describe the uncertainty in the size of the harvest, you feel that it may be described as a normal distribution with a mean value of 80,000 bushels and a standard deviation of 2,500 bushels. Find the probability that your harvest will exceed 84,000 bushels.

28. Assume that electronic microchip operating speeds are normally distributed with a mean of 2.5 gigahertz and a standard deviation of 0.4 gigahertz. What percentage of your production would you expect to be "superchips" with operating speeds of 3 gigahertz or more?

29. Although you don't know the exact total amount of payments you will receive next month, based on past experience you believe it will be approximately \$2,500 more or less than \$13,000, and will follow a normal distribution. Find the probability that you will receive between \$10,000 and \$15,000 next month.

30. A new project will be declared "successful" if you achieve a market share of 10% or more in the next two years. Your marketing department has considered all possibilities and decided that it expects the product to attain a market share of 12% in this time. However, this number is not certain. The standard deviation is forecast to be 3%, indicating the uncertainty in the 12% forecast as 3 percentage points. You may assume a normal distribution.

a.* Find the probability that the new project is successful.

b. Find the probability that the new project fails.

c. Find the probability that the new project is wildly successful, defined as achieving at least a 15% market share.

d. To assess the precision of the marketing projections, find the probability that the attained market share falls close to the projected value of 12%, that is, between 11% and 13%.

31. A manufacturing process produces semiconductor chips with a known failure rate of 6.3%. Assume that chip failures are independent of one another. You will be producing 2,000 chips tomorrow.

a. What is the name of the probability distribution of the number of defective chips produced tomorrow?

b. Find the expected number of defective chips produced.

c. Find the standard deviation of the number of defective chips.

d. Find the (approximate) probability that you will produce fewer than 130 defects.

e. Find the (approximate) probability that you will produce more than 120 defects.

f. You just learned that you will need to ship 1,860 working chips out of tomorrow's production of 2,000. What are the chances that you will succeed? Will you need to increase the scheduled number produced?

g. If you schedule 2,100 chips for production, what is the probability that you will be able to ship 1,860 working ones?

32. A union strike vote is scheduled tomorrow, and it looks close. Assume that the number of votes to strike follows a binomial distribution. You expect 300 people to vote, and you have projected a probability of 0.53 that a typical individual will vote to strike.

a. Identify n and π for this binomial random variable.

b. Find the mean and standard deviation of the number who will vote to strike.

c. Find the (approximate) probability that a strike will result (i.e., that a majority will vote to strike).

33. Reconsider the previous problem and answer each part, but assume that 1,000 people will vote. (The probability for each one remains unchanged.)

34. Assume that if you were to interview the entire population of Detroit, exactly 18.6% would say that they are ready to buy your product. You plan to interview a representative random sample of 250 people. Find the (approximate) probability that your observed sample percentage is overoptimistic, where this is defined as the observed percentage exceeding 22.5%.

35. Suppose 15% of the items in a large warehouse are defective. You have chosen a random sample of 250 items to examine in detail. Find the (approximate) probability that more than 20% of the sample is defective.

36. You are planning to interview 350 consumers randomly selected from a large list of likely sales prospects, in order to assess the value of this list and whether you should assign salespeople the task of contacting them all. Assuming that 13% of the large list will respond favorably, find (approximate) probabilities for the following:

a. More than 10% of randomly selected consumers will respond favorably.

b. More than 13% of randomly selected consumers will respond favorably.

c. More than 15% of randomly selected consumers will respond favorably.

d. Between 10% and 15% of randomly selected consumers will respond favorably.

37. You have just sent out a test mailing of a catalog to 1,000 people randomly selected from a database of 12,320 addresses. You will go ahead with the mass mailing to the remaining 11,320 addresses provided you receive orders from 2.7% or more from the test mailing within two weeks. Find the (approximate) probability that you will do the mass mailing under each of the following scenarios:

a. Assume that, in reality, exactly 2% of the population would send in an order within two weeks.

b. Assume that, in reality, exactly 3% of the population would send in an order within two weeks.

c. Assume that, in reality, exactly 4% of the population would send in an order within two weeks.

38. You expect a mean of 1,671 warranty repairs next month, with the actual outcome following a Poisson distribution.

a. Find the standard deviation of the number of such repairs.

b. Find the (approximate) probability of more than 1,700 such repairs.

39. If tomorrow is a typical day, your human resources division will expect to receive résumés from 175 job applicants. You may assume that applicants act independently of one another.
  a. What is the name of the probability distribution of the number of résumés received?
  b. What is the standard deviation of the number of résumés received?
  c. Find the approximate probability that you will receive more than 185 résumés.
  d. Find the approximate probability of a slow day, with 160 or fewer résumés received.

40. On a typical day, your clothing store takes care of 2.6 "special customers" on average. These customers are taken directly to a special room in the back, are assigned a full-time server, are given tea (or espresso) and scones, and have clothes brought to them. You may assume that the number who will arrive tomorrow follows a Poisson distribution.
  a. Find the standard deviation of the number of special customers.
  b. Find the probability that no special customers arrive tomorrow.
  c. Find the probability exactly 4 special customers will arrive tomorrow.

41. In order to earn enough to pay your firm's debt this year, you will need to be awarded at least 2 contracts. This is not usually a problem, since the yearly average is 5.1 contracts. You may assume a Poisson distribution.
  a. Find the probability that you will not earn enough to pay your firm's debt this year.
  b. Find the probability that you will be awarded exactly 3 contracts.

42. Customers arrive at random times, with an exponential distribution for the time between arrivals. Currently the mean time between customers is 6.34 minutes.
  a. Since the last customer arrived, three minutes have gone by. Find the mean time until the next customer arrives.
  b. Since the last customer arrived, 10 minutes have gone by. Find the mean time until the next customer arrives.

43. In the situation described in the previous problem, a customer has just arrived.
  a. Find the probability that the time until the arrival of the next customer is less than 3 minutes.
  b. Find the probability that the time until the arrival of the next customer is more than 10 minutes.
  c. Find the probability that the time until the arrival of the next customer is between 5 and 6 minutes.

44. A TV system is expected to last for 50,000 hours before failure. Assume an exponential distribution for the time until failure.
  a. Is the distribution skewed or symmetric?
  b. What is the standard deviation of the length of time until failure?
  c. The system has been working continuously for the past 8,500 hours and is still on. What is the expected time from now until failure? (Be careful!)

45. Assuming the appropriate probability distribution for the situation described in the preceding problem:
  a. Find the probability that the system will last 100,000 hours or more (twice the average lifetime).
  b. The system is guaranteed to last at least 5,000 hours. What percentage of production is expected to fail during the guarantee period?

46. Compare the "probability of being within one standard deviation of the mean" for the exponential and normal distributions.

Database Exercises

Problems marked with an asterisk (*) are solved in the Self Test in Appendix C.

Refer to the employee database in Appendix A.

1. View each column as a collection of independent observations of a random variable.
  a. In each case, what kind of variable is represented, continuous or discrete? Why?
  b.* Consider the event "annual salary is above \$40,000." Find the value of the binomial random variable X that represents the number of times this event occurred. Also find the binomial proportion p and say what it represents.
  c. What fraction of employees are male? Interpret this number as a binomial proportion. What is n?

2. You have a position open and are trying to hire a new person. Assume that the new person's experience will follow a normal distribution with the mean and (sample) standard deviation of your current employees.
  a. Find the probability that the new person will have more than six years of experience.
  b. Find the probability that the new person will have less than three years of experience.
  c. Find the probability that the new person will have between four and seven years of experience.

3. Suppose males and females are equally likely and that the number of each gender follows a binomial distribution. (Note that the database contains observations of random variables, not the random variables themselves.)
  a. Find n and π for the binomial distribution of the number of males.
  b. Find n and π for the binomial distribution of the number of females.
  c. Find the observed value of X for the number of females.
  d. Use the normal approximation to the binomial distribution to find the probability of observing this many females (your answer to part c) or fewer in the database.

Projects

1. Choose a continuous random quantity that you might deal with in your current or future work as an executive. Model it as a normally distributed random variable and estimate (i.e., guess) the mean and standard deviation. Identify three events of interest to you relating to this random variable and compute their probabilities. Briefly discuss what you have learned.

2. Choose a discrete random quantity (taking on from 3 to 10 different values) that you might deal with in your current or future work as an executive. Estimate (i.e., guess) the probability distribution. Compute the mean and standard deviation. Identify two events of interest to you relating to this random variable and compute their probabilities. Briefly discuss what you have learned.
3. Choose a binomial random quantity that you might deal with in your current or future work as an executive. Estimate (i.e., guess) the value of n and π. Compute the mean and standard deviation. Identify two events of interest to you relating to this random variable and compute their probabilities. Briefly discuss what you have learned.
4. On the Internet, find and record observations on at least five different random variables such as stock market indices, interest rates, corporate sales, or any business-related topic of interest to you.

Case

The Option Value of an Oil Lease

There's an oil leasing opportunity that looks too good to be true, and it probably *is* too good to be true: An estimated 1,500,000 barrels of oil sitting underground that can be leased for three years for just $1,300,000. It looks like a golden opportunity: Pay just over a million, bring the oil to the surface, sell it at the current spot price of $76.45 per barrel, and retire.

However, upon closer investigation, you come across the facts that explain why nobody else has snapped up this "opportunity." Evidently, it is difficult to remove the oil from the ground due to the geology and the remote location. A careful analysis shows that estimated costs of extracting the oil are a whopping $120,000,000. You conclude that by developing this oil field, you would actually *lose* money. Oh well.

During the next week, although you are busy investigating other capital investment opportunities, your thoughts keep returning to this particular project. In particular, the fact that the lease is so cheap and that it lasts for three years inspires you to do a *What if* scenario analysis, recognizing that there is no obligation to extract the oil and that it could be extracted fairly quickly (taking a few months) at any time during the three-year lease. You are wondering: What if the price of oil rises enough during the three years for it to be profitable to develop the oil field? If so, then you would extract the oil. But if the price of oil didn't rise enough, you would let the term of the lease expire in three years, leaving the oil still in the ground. You would let the future price of oil determine whether or not to exercise the option to extract the oil.

But such a proposition is risky! How much risk? What are the potential rewards? You have identified the following basic probability structure for the source of uncertainty in this situation:

Future Price of Oil	Probability
60	0.10
70	0.15
80	0.20
90	0.30
100	0.15
110	0.10

Discussion Questions

1. How much money would you make if there were no costs of extraction? Would this be enough to retire?
2. Would you indeed lose money if you leased and extracted immediately, considering the costs of extraction? How much money?
3. Continue the scenario analysis by computing the future net payoff implied by each of the future prices of oil. To do this, multiply the price of oil by the number of barrels; then subtract the cost of extraction. If this is negative, you simply won't develop the field, so change negative values to zero. (At this point, do not subtract the lease cost, because we are assuming that it has already been paid.)
4. Find the average future net payoff, less the cost of the lease. How much, on average, would you gain (or lose) by leasing this oil field? (You may ignore the time value of money.)
5. How risky is this proposition?
6. Should you lease or not?