Title: Business Analytics Data Analysis and Decision Making

Author: Albright, Winston

Article Title: Introduction to Simulation Modeling

Volume:

Issue:

Year: 2017

Pages : 759-828

Class: BUDT732

Semester: Fall 2018

Instructor: Bardossy, Maria

6TH EDITION

Business Analytics: Data Analysis and Decision Making

S. Christian Albright
Kelly School of Business, Indiana University, Emeritus

Wayne L. Winston
Kelly School of Business, Indiana University

CENGAGE Learning

Australia • Brazil • Mexico • Singapore • United Kingdom • United States

CENGAGE Learning

Business Analytics: Data Analysis & Decision Making, Sixth Edition

S. Christian Albright and Wayne L. Winston

Vice President, General Manager, Social Science & Qualitative Business: Balraj Kalsi

Product Director: Mike Schenk

Product Manager: Aaron Arnsparger

Associate Content Developer: Brad Sullender

Product Assistant: Audrey Jacobs

Sr. Marketing Manager: Nathan Anderson

Manufacturing Planner: Ron J. Montgomery

Art and Cover Direction, Production Management, and Composition: Lumina Datamatics, Inc.

Cover Image: iStockphoto.com/derrrek

Intellectual Property

Analyst: Brittani Morgan

Project Manager: Nick Barrows

For product information and technology assistance, contact us at
Cengage Learning Customer & Sales Support, 1-800-354-9706

For permission to use material from this text or product,
submit all requests online at **www.cengage.com/permissions**
Further permissions questions can be emailed to
permissionrequest@cengage.com

Library of Congress Control Number: 2016934323

ISBN: 978-1-305-94754-2

Loose Leaf Edition ISBN: 978-1-337-09643-0

Cengage Learning
20 Channel Center Street
Boston, MA 02210
USA

Cengage Learning is a leading provider of customized learning solutions with employees residing in nearly 40 different countries and sales in more than 125 countries around the world. Find your local representative at **www.cengage.com.**

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

To learn more about Cengage Learning Solutions, visit **www.cengage.com**

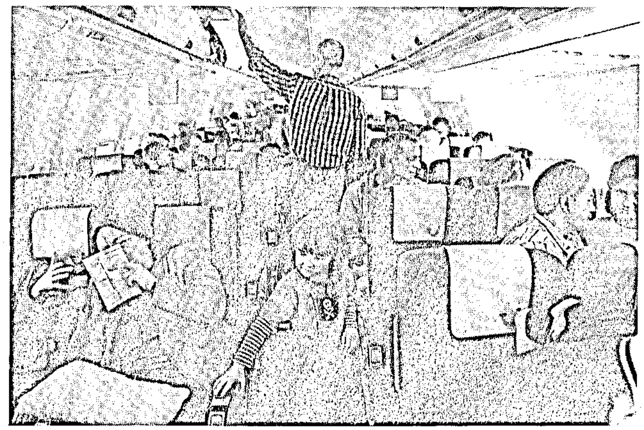
Purchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com**

Printed in the United States of America
Print Number: 04 Print Year: 2018

CHAPTER

15

Introduction to Simulation Modeling


Caro/Alamy

DEVELOPING BOARDING STRATEGIES AT AMERICA WEST

Management science often attempts to solve problems that we all experience. One such problem is the boarding process for airline flights. As customers, we all hate to wait while travelers boarding ahead of us store their luggage and block the aisles. But this is also a big problem for the airlines. Airlines lose money when their airplanes are on the ground, so they have a real incentive to reduce the turnaround time from when a plane lands until it departs on its next flight. Of course, the turnaround time is influenced by several factors, including passenger deplaning, baggage unloading, fueling, cargo unloading, airplane maintenance, cargo loading, baggage loading, and passenger boarding. Airlines try to perform all of these tasks as efficiently as possible, but passenger boarding is particularly difficult to shorten. Although the airlines want passengers to board as quickly as possible, they don't want to use measures that might antagonize their passengers.

One study by van den Briel et al. (2005) indicates how a combination of management science methods, including simulation, was used to make passenger boarding more efficient at America West Airlines. America West (which merged with US Airways in 2006) was a major U.S. carrier based in Phoenix, Arizona. It served more destinations nonstop than any other airline. The airline's fleet consisted of Airbus A320s, Airbus A319s, Boeing 757s, Boeing 737s, and Airbus A318s.

At the time of the study, airlines used a variety of boarding strategies, but the predominant strategy was the back-to-front (BF) strategy where, after boarding first-class passengers and passengers with special needs, the rest of the passengers are boarded in groups, starting with rows in the back of the plane. As the authors suspected (and most of us have experienced), this strategy still results in significant congestion. Within a given section of the plane (the back, say), passengers storing luggage in overhead compartments can block an aisle. Also, people in the aisle or middle seat often need to get back into the aisle to let window-seat passengers be seated. The authors developed an integer programming (IP) model to minimize the number of such aisle blockages. The decision variables determined which groups of seats should be boarded in which order. Of course, the BF strategy was one possible feasible solution, but it turned out to be a suboptimal solution. The IP model suggested that the best solution was an outside-in (OI) strategy, where groups of passengers in window seats board first, then groups in the middle seats, and finally groups in aisle seats, with all of these groups going essentially in a back-to-front order.

The authors recognized that their IP model was at best an idealized model of how passengers actually behave. Its biggest drawback is that it ignores the inherent randomness in passenger behavior. Therefore, they followed up their optimization model with a simulation model. As they state, "We used simulation to validate the analytical model and to obtain a finer level of detail." This validation of an approximate or idealized analytical model is a common use for simulation. To make the simulation as realistic as possible, they used two cameras, one inside the plane and one inside the bridge leading to the plane, to tape customer behavior. By analyzing the tapes, they were able to estimate the required inputs to their simulation model, such as the time between passengers, walking speed, blocking time, and time to store luggage in overhead compartments. After the basic simulation model was developed, it was used as a tool to evaluate various boarding strategies suggested by the IP model. It also allowed the authors to experiment with changes to the overall boarding process that might be beneficial. For example, reducing congestion *inside* the airplane is not very helpful if the gate agent at the entrance to the bridge processes passengers too slowly. Their final recommendation, based on a series of simulation experiments, was to add a second gate agent (there had been only one before) and to board passengers in six groups using an OI strategy. The simulation model suggested that this could reduce the boarding time by about 37%.

The authors' recommendations were implemented first as a pilot project and then systemwide. The pilot results were impressive, with a 39% reduction in boarding times. By September 2003, the new boarding strategies had been implemented in 80% of America West's airports, with a decrease in departure delays as much as 60.1%. Besides this obvious benefit to the airline, customers also appear to be happier. Now they can easily understand when to queue up for boarding, and they experience less blocking after they get inside the plane. ◘

15-1 INTRODUCTION

A **simulation model** is a computer model that imitates a real-life situation. It is like other mathematical models, but it explicitly incorporates uncertainty in one or more input variables. When you run a simulation, you allow these random input variables to take on various values, and you keep track of any resulting output variables of interest. In this way, you are able to see how the outputs vary as a function of the varying inputs.

The fundamental advantage of a simulation model is that it provides an entire distribution of results, not simply a single bottom-line result. As an example, suppose an automobile manufacturer is planning to develop and market a new model car. The company is ultimately interested in the net present value (NPV) of the cash flows from this car over the next 10 years. However, there are many uncertainties surrounding this car, including the yearly customer demands for it, the cost of developing it, and others. The company could develop a spreadsheet model for the 10-year NPV, using its *best guesses* for these uncertain quantities. It could then report the NPV based on these best guesses. However, this analysis would be incomplete and probably misleading—there is no guarantee that the NPV based on best-guess inputs is representative of the NPV that will actually occur. It is much better to treat the uncertainty explicitly with a simulation model. This involves entering probability distributions for the uncertain quantities and seeing how the NPV varies as the uncertain quantities vary.

Each different set of values for the uncertain quantities is a scenario. Simulation allows the company to generate many scenarios, each leading to a particular NPV. In the end, it sees a whole distribution of NPVs, not a single best guess. The company can see what the NPV will be on average, and it can also see worst-case and best-case results.

These approaches are summarized in Figures 15.1 and 15.2. Figure 15.1 indicates that the deterministic (nonsimulation) approach, using best guesses for the uncertain inputs, is generally *not* the appropriate method. It leads to the "flaw of averages," as we will discuss later in the chapter. The problem is that the outputs from the deterministic model are often not representative of the true outputs. The appropriate method is shown in Figure 15.2. Here the uncertainty is modeled explicitly with random inputs, and the end result is a probability distribution for each of the important outputs.

Figure 15.1
Inappropriate Deterministic Model

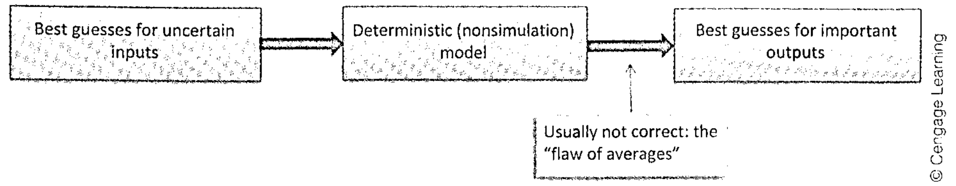
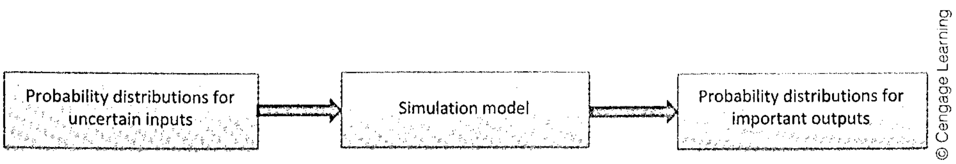


Figure 15.2
Appropriate Simulation Model



Simulation models are also useful for determining how sensitive a system is to changes in operating conditions. For example, the operations of a supermarket could be simulated. Once the simulation model has been developed, it could then be run (with suitable modifications) to ask a number of what-if questions. For example, if the supermarket experiences a 20% increase in business, what will happen to the average time customers must wait for service?

A huge benefit of computer simulation is that it enables managers to answer these types of what-if questions without actually changing (or building) a physical system. For example, the supermarket might want to experiment with the number of open registers to see the effect on customer waiting times. The only way it can *physically* experiment with more registers than it currently owns is to purchase more equipment. Then if it determines that this equipment is not a good investment—customer waiting times do not decrease appreciably—the company is stuck with expensive equipment it doesn't need. Computer simulation is a much less expensive alternative. It provides the company with an electronic replica of what would happen *if* the new equipment were purchased. Then, if the simulation indicates that the new equipment is worth the cost, the company can be confident that purchasing it is the right decision. Otherwise, it can abandon the idea of the new equipment *before* the equipment has been purchased.

Spreadsheet simulation modeling is quite similar to the other modeling applications in this book. You begin with input variables and then relate these with appropriate Excel® formulas to produce output variables of interest. The main difference is that simulation uses *random* numbers to drive the whole process. These random numbers are generated with special functions that we will discuss in detail. Each time the spreadsheet recalculates, all of the random numbers change. This provides the ability to model the logical process once and then use Excel's recalculation ability to generate many different scenarios. By collecting the data from these scenarios, you can see the most likely values of the outputs and the best-case and worst-case values of the outputs.

In this chapter we begin by illustrating spreadsheet models that can be developed with built-in Excel functionality. However, because simulation is such an important tool for analyzing real problems, add-ins to Excel have been developed to streamline the process of developing and analyzing simulation models. Therefore, we then introduce @RISK, one of the most popular simulation add-ins. This add-in not only augments the simulation capabilities of Excel, but it also enables you to analyze models much more quickly and easily.

The purpose of this chapter is to introduce basic simulation concepts, show how simulation models can be developed in Excel, and demonstrate the capabilities of @RISK. Then in the next chapter, armed with the necessary simulation tools, we will explore a variety of simulation models.

Before proceeding, you might ask whether simulation is really used in the business world. The answer is a resounding "yes." The chapter opener described an airline example, and many other examples can be found online. For example, if you visit www.palisade.com, you will see descriptions of interesting @RISK applications from companies that regularly use this add-in. Simulation has always been a powerful tool, but until the introduction of Excel add-ins such as @RISK, it had limited use for several reasons. It typically required specialized software that was either expensive and difficult to learn, or it required tedious computer programming. Fortunately, in the past two decades, spreadsheet simulation, together with Excel add-ins such as @RISK, has put this powerful methodology in the hands of the masses—people like you and the companies you are likely to work for. Many businesses now understand that there is no longer any reason to ignore uncertainty; they can model it directly with spreadsheet simulation.

15-2 PROBABILITY DISTRIBUTIONS FOR INPUT VARIABLES

In spreadsheet simulation models, input cells can contain random numbers. Any output cells then vary as these random inputs change.

In this section we discuss the building blocks of spreadsheet simulation models: **probability distributions for input variables** that capture uncertainty. All spreadsheet simulation models are similar to the spreadsheet models from previous chapters. They have a number of cells that contain values of input variables. The other cells then contain formulas that embed the logic of the model and eventually lead to the output variable(s) of interest. The primary difference between the spreadsheet models you have developed so far and simulation models is that at least one of the input variable cells in a simulation model contains *random* numbers. Each time the spreadsheet recalculates, the random numbers change, and the new random values of the inputs produce new values of the outputs. This is the essence of simulation—it enables you to see how outputs vary as random inputs change.

Excel Tip: Recalculation Key
The easiest way to make a spreadsheet recalculate is to press the **F9 key**. *This is often called the "recalc" key.*

Technically speaking, input cells do not contain random numbers; they contain *probability distributions.* In general, a probability distribution indicates the possible values of a variable and the probabilities of these values. As a very simple example, you might indicate by an appropriate formula (to be described later) that you want a probability distribution with possible values 50 and 100, and corresponding probabilities 0.7 and 0.3. If you force the sheet to recalculate repeatedly and watch this input cell, you will see the value 50 about 70% of the time and the value 100 about 30% of the time. No other values besides 50 and 100 will appear.

When you enter a given probability distribution in a random input cell, you are describing the possible values and the probabilities of these values that you believe mirror reality. There are many probability distributions to choose from, and you should always attempt to choose an *appropriate* distribution for each specific problem. This is not necessarily an easy task. Therefore, we address it in this section by answering several key questions:

- What types of probability distributions are available, and why do you choose one probability distribution rather than another in any particular simulation model?
- Which probability distributions can you use in simulation models, and how do you invoke them with Excel formulas?

In later sections we address one additional question: Does the choice of input probability distribution really matter—that is, are the *outputs* from the simulation sensitive to this choice?

FUNDAMENTAL INSIGHT

Basic Elements of Spreadsheet Simulation

A spreadsheet simulation model requires three elements: (1) a method for entering random quantities from specified probability distributions in input cells, (2) the usual types of Excel formulas for relating outputs to inputs, and (3) the ability to make the spreadsheet recalculate many times and capture the resulting outputs for statistical analysis. Excel has some capabilities for performing these steps, but Excel add-ins such as @RISK provide excellent tools for automating the process.

15-2a Types of Probability Distributions

Imagine a toolbox that contains the probability distributions you know and understand. As you obtain more experience in simulation modeling, you will naturally add probability distributions to your toolbox that you can then use in future simulation models. We begin by adding a few useful probability distributions to this toolbox. However, before adding any specific distributions, it is useful to provide a brief review of some important general characteristics of probability distributions.[1] These include the following distinctions:

- Discrete versus continuous
- Symmetric versus skewed
- Bounded versus unbounded
- Nonnegative versus unrestricted

FUNDAMENTAL INSIGHT

Choosing Probability Distributions for Uncertain Inputs

In simulation models, it is important to choose *appropriate* probability distributions for all uncertain inputs. These choices can strongly affect the results. Unfortunately, there are no "right answers." You need to choose the probability distributions that best describe your uncertainty, and this is not necessarily easy. However, the properties discussed in this section provide you with useful guidelines for making reasonable choices.

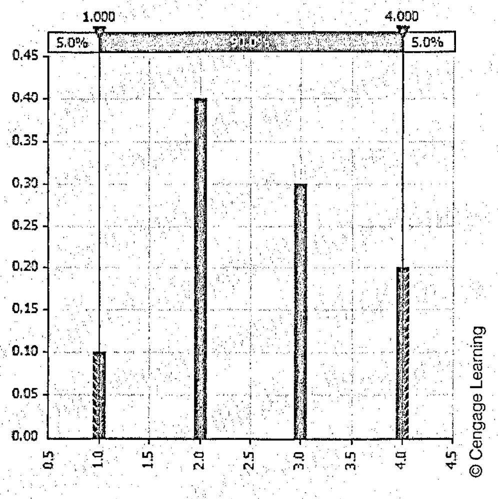
[1]Most of this material was covered in Chapters 2, 4, and 5, but it is reviewed here.

Discrete Versus Continuous

A probability distribution is *discrete* if it has a finite number of possible values.[2] For example, if you throw two dice and look at the sum of the faces showing, there are only 11 discrete possibilities: the integers 2 through 12. In contrast, a probability distribution is *continuous* if its possible values are essentially a continuum. An example is the amount of rain that falls during a month in Indiana. It could be any decimal value from 0 to, say, 15 inches.

The graph of a discrete distribution is a series of spikes, as shown in Figure 15.3.[3] The height of each spike is the probability of the corresponding value.

Figure 15.3
A Typical Discrete Probability Distribution



The heights above a density function are not probabilities, but they still indicate relative likelihoods of the possible values.

In contrast, a continuous distribution is characterized by a *density function*, a smooth curve as shown in Figure 15.4. Recall from Chapter 5 that the height of the density function above any value indicates the relative likelihood of that value, and probabilities can be calculated as areas under the curve.

Sometimes it is convenient to treat a discrete probability distribution as continuous, and vice versa. For example, consider a student's random score on an exam that has 1000 possible points. If the grader scores each exam to the nearest integer, then even though the score is really discrete with many possible integer values, it is probably more convenient to model its distribution as a continuum. Continuous probability distributions are typically more intuitive and easier to work with than discrete distributions in cases such as this, where there are many possible values. In contrast, continuous distributions are sometimes *discretized* for simplicity. In this case, the continuum of possible values is replaced by a few typical values.

Symmetric Versus Skewed

A probability distribution can either be symmetric or skewed to the left or right. Figures 15.4, 15.5, and 15.6 provide examples of each of these. You typically choose between a symmetric and skewed distribution on the basis of realism. For example, if you want to model a student's score on a 100-point exam, you will probably choose a left-skewed distribution. This is because a few poorly prepared students typically "pull down the curve." On the other hand, if you want to model the time it takes to serve a customer at a bank, you will probably choose a right-skewed distribution. This is because most customers take only a minute or two, but a few

[2]Actually, it is possible for a discrete variable to have a *countably infinite* number of possible values, such as all the nonnegative integers 0, 1, 2, and so on. However, this is not an important distinction for practical applications.
[3]This figure and several later figures have been captured from Palisade's @RISK add-in.

Figure 15.4
A Typical Continuous Probability Distribution

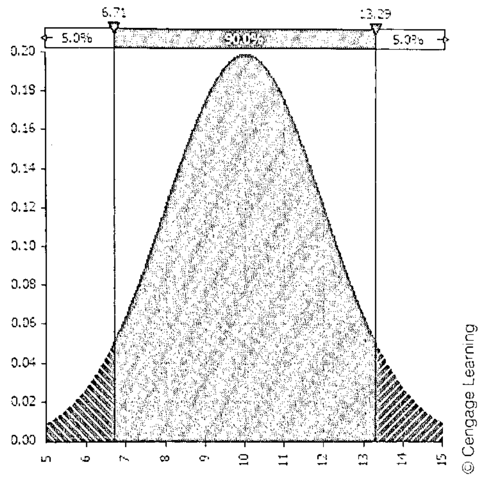


Figure 15.5
A Positively Skewed Probability Distribution

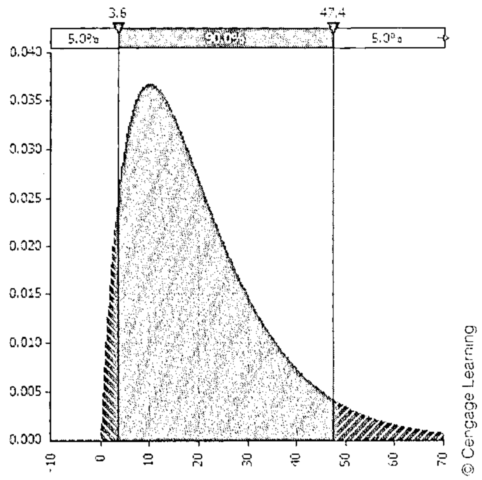
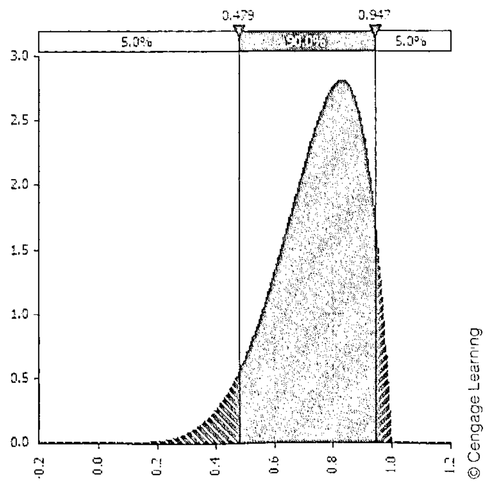


Figure 15.6
A Negatively Skewed Probability Distribution



customers take a long time. Finally, if you want to model the monthly return on a stock, you might choose a distribution symmetric around zero, reasoning that the stock return is just as likely to be positive as negative and there is no obvious reason for skewness in either direction.

Bounded Versus Unbounded

A probability distribution is *bounded* if there are values *A* and *B* such that no possible value can be less than *A* or greater than *B*. The value *A* is then the *minimum* possible value, and the value *B* is the *maximum* possible value. The distribution is *unbounded* if there are no such bounds. Actually, it is possible for a distribution to be bounded in one direction but not the other. As an example, the distribution of scores on a 100-point exam is bounded between 0 and 100. In contrast, the distribution of the amount of damages Mr. Jones submits to his insurance company in a year is bounded on the left by 0, but there is no natural upper bound. Therefore, you might model this amount with a distribution that is bounded by 0 on the left but is unbounded on the right. Alternatively, if you believe that no damage amount larger than $20,000 can occur, you could model this amount with a distribution that is bounded in both directions.

Nonnegative versus Unrestricted

One important special case of bounded distributions is when the only possible values are *nonnegative*. For example, if you want to model the random cost of manufacturing a new product, you know for sure that this cost must be nonnegative. There are many other such examples. In such cases, you should model the randomness with a probability distribution that is bounded below by 0. This rules out negative values that make no practical sense.

15-2b Common Probability Distributions

Think of the ***Probability Distributions.xlsx*** *file as a "dictionary" of the most commonly used distributions. Keep it handy for reference.*

Now that you know the *types* of probability distributions available, you can add some common probability distributions to your toolbox. The file **Probability Distributions.xlsx** was developed to help you learn and explore these. Each sheet in this file illustrates a particular probability distribution. It describes the general characteristics of the distribution, indicates how you can generate random numbers from the distribution either with Excel's built-in functions or with @RISK functions, and it includes histograms of these distributions from simulated data to illustrate their shapes.[4]

A family of distributions has a common name, such as "normal." Each member of the family is specified by one or more numerical parameters.

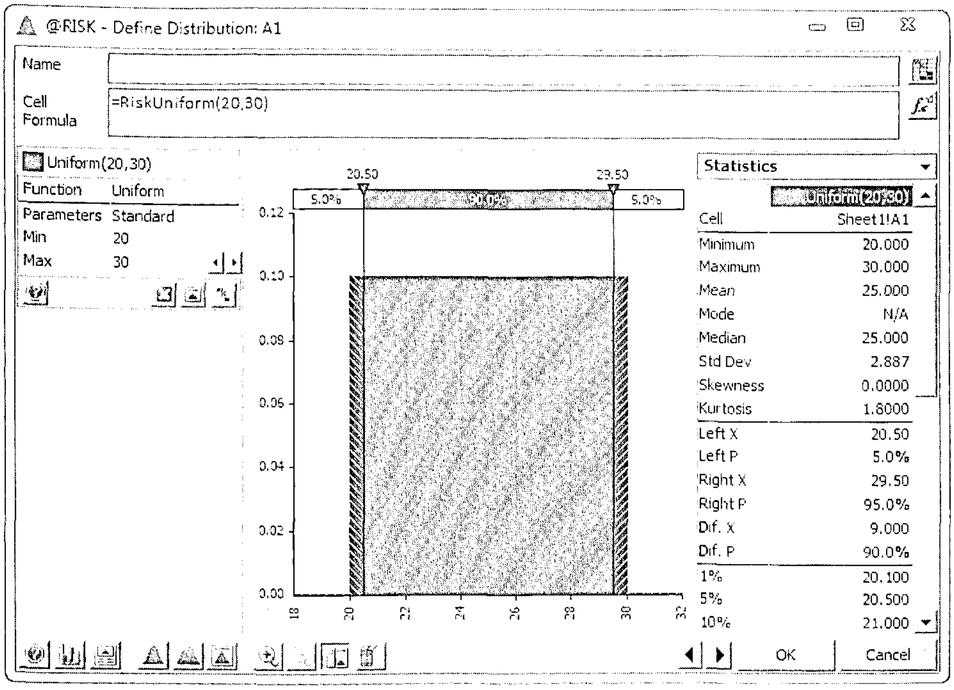
It is important to realize that each of the following distributions is really a *family* of distributions. Each member of the family is specified by one or more parameters. For example, there is not a *single* normal distribution; there is a normal distribution for each possible mean and standard deviation you specify. Therefore, when you try to find an appropriate input probability distribution in a simulation model, you first have to choose an appropriate family, and then you have to select the appropriate parameters for that family.

Uniform Distribution

The **uniform distribution** is the "flat" distribution illustrated in Figure 15.7. It is bounded by a minimum and a maximum, and all values between these two extremes are equally likely. You can think of this as the "I have no idea" distribution. For example, a manager might realize that a building cost is uncertain. If she can state only that, "I know the cost will be between $20,000 and $30,000, but other than this, I have no idea what the cost will be," then a uniform distribution from $20,000 to $30,000 is a natural choice. However, even though some people do sometimes use the uniform distribution in such situations, the

[4]In later sections of this chapter, and all through the next chapter, we discuss much of @RISK's functionality. For this section, the only functionality we use is @RISK's collection of functions, such as RISKNORMAL and RISKTRIANG, for generating random numbers from various probability distributions. You can skim the details of these functions for now and refer back to them as necessary in later sections.

Figure 15.7
Uniform Distribution



uniform is usually not very realistic. If the manager really thinks about it, she can probably provide more information about the uncertain cost, such as, "The cost is more likely to be close to $25,000 than to either of the extremes." Then some distribution other than the uniform is more appropriate.

Regardless of whether the uniform distribution is an appropriate candidate as an input distribution, it is important for another reason. All simulation software packages, including Excel, are capable of generating random numbers uniformly distributed between 0 and 1. These are the building blocks of most simulated random numbers, in that random numbers from other probability distributions are generated from them.

The RAND function is Excel's "building block" function for generating random numbers.

In Excel, you can generate a random number between 0 and 1 by entering the formula

=RAND()

in any cell. (The parentheses to the right of RAND indicate that this is an Excel function with no arguments. These parentheses must be included.)

RAND and RANDBETWEEN Functions

Excel Function: RAND

To generate a random number equally likely to be anywhere between 0 and 1, enter the formula ***=RAND()*** *into any cell. Press the F9 key, or recalculate in any other way, to make it change randomly.*

In addition to being between 0 and 1, the numbers created by this function have two important properties.

- **Uniform property.** Each time you enter the RAND function in a cell, all numbers between 0 and 1 have the same chance of occurring. This means that approximately 10% of the numbers generated by the RAND function will be between 0.0 and 0.1; 10% of the numbers will be between 0.65 and 0.75; 60% of the numbers will be between 0.20 and 0.80; and so on. This property explains why the random numbers are said to be *uniformly distributed* between 0 and 1.

- **Independence property.** Different random numbers generated by RAND functions are *probabilistically independent*. This implies that when you generate a random number in cell A5, say, it has no effect on the values of any other random numbers generated in the spreadsheet. For example, if one call to the RAND function yields a large random number such as 0.98, there is no reason to suspect that the next call to RAND will yield an abnormally small (or large) random number; it is unaffected by the value of the first random number.

Excel Function: RANDBETWEEN

*Besides the RAND function, there is one other function built into Excel that generates random numbers, the RANDBETWEEN function. It takes two integer arguments, as in **=RANDBETWEEN(1,6)**, and returns a random integer between these values (including the endpoints) so that all such integers are equally likely.*

To illustrate the RAND function, open a new workbook, enter the formula **=RAND()** in cell A4, and copy it to the range A4:A503. This generates 500 random numbers. Figure 15.8 displays a possible set of values. However, when you try this on your PC, you

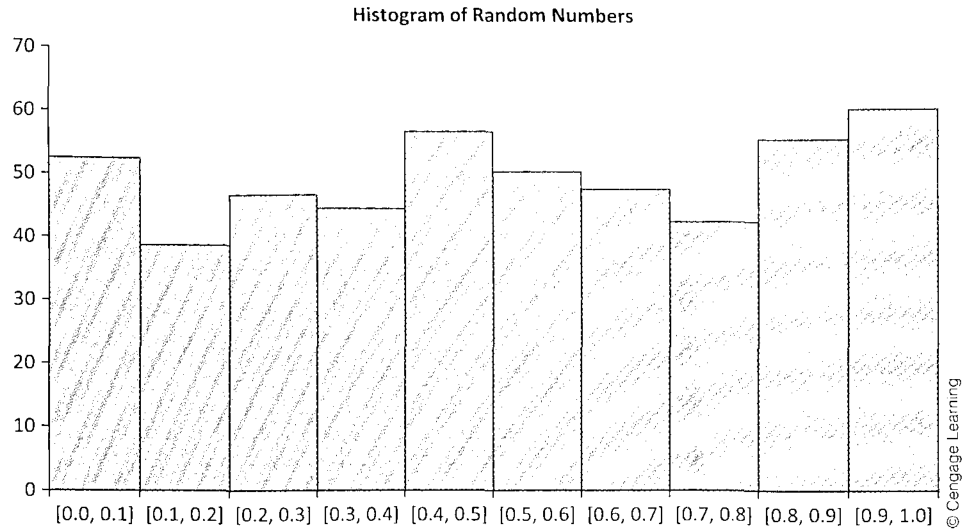
Figure 15.8
Uniformly Distributed Random Numbers Generated by the RAND Function

	A	B	C
1	**500 random numbers from RAND function**		
2			
3	Random #		
4	0.429866189		
5	0.96337384		
6	0.411422456		
7	0.909329143		
8	0.54307235		
9	0.75894605		
10	0.920756879		
500	0.720597182		
501	0.054328298		
502	0.265368444		
503	0.402328268		

will undoubtedly obtain *different* random numbers. This is an inherent characteristic of simulation—no two answers are ever exactly alike. Now press the recalc (F9) key. All of the random numbers will change. In fact, each time you press the F9 key or do anything to make your spreadsheet recalculate, all of the cells containing the RAND function will change.

A histogram of the 500 random numbers appears in Figure 15.9. (Again, if you try this on your PC, the shape of your histogram will not be identical to the one shown in Figure 15.9, because it will be based on *different* random numbers.) From property 1, you would expect *equal* numbers of observations in the 10 categories. Obviously, the heights of the bars are *not* exactly equal, but the differences are due to chance—not to a faulty random number generator.

Figure 15.9
Histogram of the 500 Random Numbers Generated by the RAND Function



Technical Note: Pseudo-random Numbers

The "random" numbers generated by the RAND function (or by the random number generator in any simulation software package) are not really random. They are sometimes called pseudo-random numbers. *Each successive random number follows the previous random number by a complex arithmetic operation. If you happen to know the details of this arithmetic operation, you can predict ahead of time exactly which random numbers will be generated by the RAND function. This is quite different from using a "true" random mechanism, such as spinning a wheel, to get the next random number—a mechanism that would be impractical to implement on a computer. Mathematicians and computer scientists have studied many ways to produce random numbers that have the two properties we just discussed, and they have developed many competing random number generators such as the RAND function in Excel. The technical details are not important here. The important point is that these random number generators produce numbers that appear to be random and are useful for simulation modeling.*

It is simple to generate a uniformly distributed random number with a minimum and maximum other than 0 and 1. For example, the formula

=200+100*RAND()

generates a number uniformly distributed between 200 and 300. (Make sure you see why.) Alternatively, you can use the @RISK formula[5]

=RISKUNIFORM(200,300)

You can take a look at this and other properties of the uniform distribution on the Uniform sheet in the **Probability Distributions.xlsx** file. (See Figure 15.10.)

@RISK Function: RISKUNIFORM

To generate a random number from any uniform distribution, enter the formula ***=RISKUNIFORM(MinVal,MaxVal)*** *in any cell. Here, MinVal and MaxVal are the minimum and maximum possible values. Note that if MinVal is 0 and MaxVal is 1, this function is equivalent to Excel's RAND function.*

[5]As we have done with other Excel functions, we capitalize the @RISK functions, such as RISKUNIFORM, in the text. However, this is not necessary when you enter the formulas in Excel.

Figure 15.10
Properties of Uniform Distribution

	A	B	C	D	E	F	G
1	**Uniform distribution**						
2							
3	**Characteristics**						
4	Continuous						
5	Symmetric						
6	Bounded in both directions						
7	Not necessarily positive (depends on bounds)						
8							
9	**Parameters**						
10	MinVal	50					
11	MaxVal	100					
12							
13	**Excel**		**Example**				
14	=MinVal + (MaxVal-MinVal)*RAND()		92.310980				
15							
16	**@RISK**						
17	=RISKUNIFORM(MinVal,MaxVal)		77.276447				

This is a flat distribution between two values, labeled here MinVal and MaxVal. Note that if MinVal=0 and MaxVal=1, then you can just use Excel's RAND function.

Freezing Random Numbers

The automatic recalculation of random numbers can be useful sometimes and annoying at other times. There are situations when you want the random numbers to stay fixed—that is, you want to **freeze random numbers** at their current values. The following three-step method does this.

1. Select the range that you want to freeze, such as A4:A503 in Figure 15.8.
2. Press Ctrl+c to copy this range.
3. With the same range still selected, select the Paste Values option from the Paste drop-down menu on the Home ribbon. This procedure pastes a copy of the range onto itself, except that the entries are now numbers, not formulas. Therefore, whenever the spreadsheet recalculates, these numbers do not change.

Random numbers that have been frozen do not change when you press the F9 key.

Each sheet in the **Probability Distributions.xlsx** file has a list of 500 random numbers that have been frozen. The histograms in the sheets are based on the frozen random numbers. However, we encourage you to enter "live" random numbers in column B over the frozen ones and see how the histogram changes when you press F9.

15-2c Using @RISK to Explore Probability Distributions

Exploring Distributions with @RISK

The **Probability Distributions.xlsx** file illustrates a few frequently used probability distributions, and it shows the formulas required to generate random numbers from these distributions. Another option is to use Palisade's @RISK add-in, which allows you to experiment with probability distributions with its **distribution functions**. Essentially, it allows you to see the shapes of various distributions and to calculate probabilities for them, all in a user-friendly graphical interface.

To run @RISK, click the Windows Start button, go to the Programs tab, locate the Palisades DecisionTools® Suite, and select @RISK. After a few seconds, you will see the welcome screen, which you can close. At this point, you should have an @RISK tab and corresponding ribbon. (You will also see a Project button in the Tools group if you have Microsoft Project installed on your computer.) Select a blank cell in your worksheet, and then click the Define Distributions button on the @RISK ribbon (see Figure 15.11). You will see one of several galleries of distributions, depending on the tab you select. For example, Figure 15.12 shows the gallery of common distributions. Highlight one of the distributions and click Select Distribution. For example, choose the uniform distribution and enter 75 and 150 as the Min and Max parameters. You will see the shape of the distribution and a list of summary measures to the right, as shown in Figure 15.13. For example, it indicates that the mean and standard deviation of this uniform distribution are 112.5 and 21.65.

Figure 15.11 @RISK Ribbon

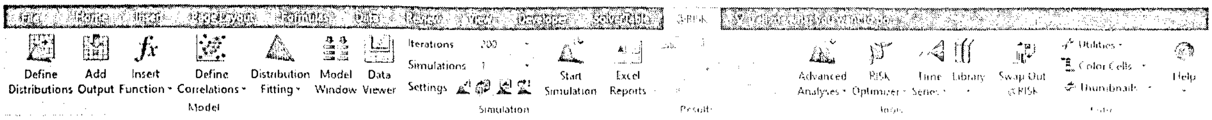
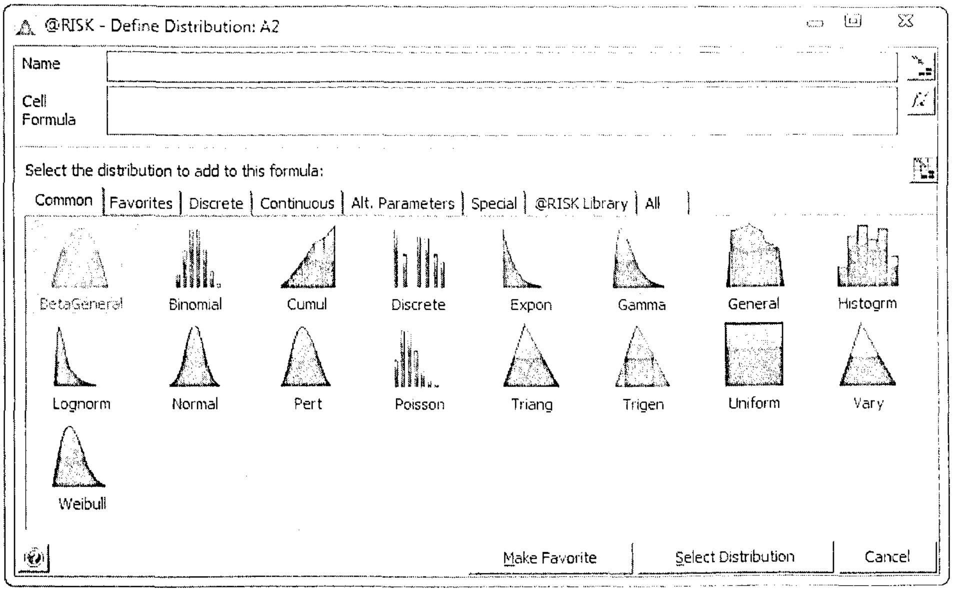


Figure 15.12 Gallery of Common Distributions

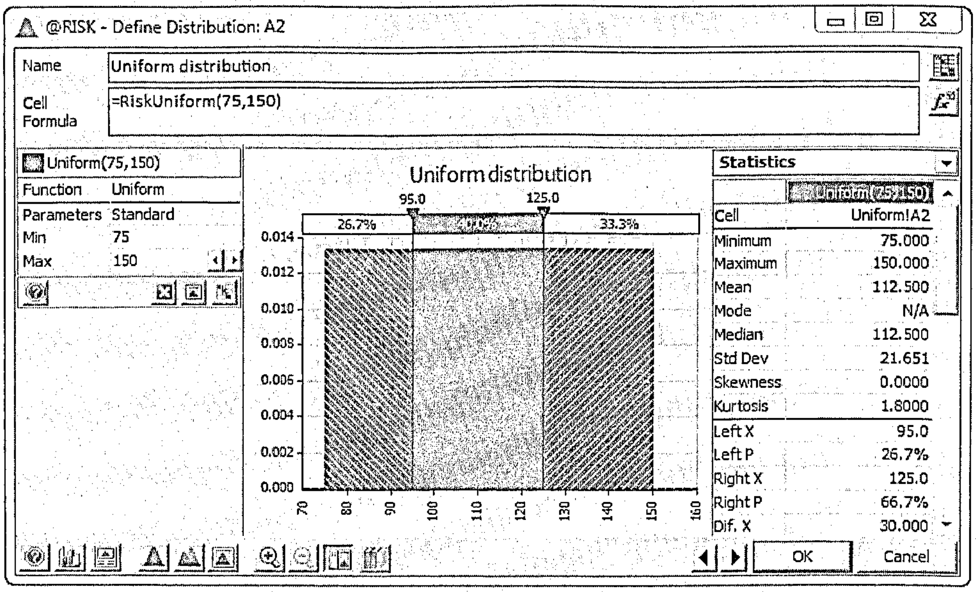


Everything in this window is interactive. Suppose you want to find the probability that a value from this distribution is less than 95. You can drag the left-hand "slider" in the diagram (the vertical line with the triangle at the top) to the position 95, as shown in Figure 15.13. You see immediately that the left-hand probability is 0.267. Similarly, if you want the probability that a value from this distribution is greater than 125, you can drag the right-hand slider to the position 125 to see that the required probability is 0.3333. (Rather than sliding, you can enter the numbers, such as 95 and 125, directly into the areas above the sliders.)

The interactive capabilities of @RISK's Define Distribution window, with its sliders, make it perfect for finding probabilities or percentiles for any given distribution.

You can also enter probabilities instead of values. For example, if you want the value such that there is probability 0.10 to the left of it—the 10th percentile—you can enter 10% in the left space above the chart. You will see that the corresponding value is 82.5. Similarly, if you want the value such that there is probability 0.10 to the right of it, you can enter 10% in the right space above the chart, and you will see that the corresponding value is 142.5.

Figure 15.13
Uniform Distribution (from @RISK)



The Define Distribution window in @RISK is quick and easy. We urge you to use it and experiment with some of its options. By the way, you can click the third button from the left at the bottom of the window to copy the chart into an Excel worksheet. However, you then lose the interactive capabilities, such as moving the sliders.

Discrete Distribution

A **discrete distribution** is useful for many situations, either when the uncertain quantity is not really continuous (the number of televisions demanded, for example) or when you want a discrete approximation to a continuous variable. All you need to specify are the possible values and their probabilities, making sure that the probabilities sum to 1. Because of this flexibility in specifying values and probabilities, discrete distributions can have practically any shape.

@RISK's way of generating a discrete random number is much simpler and more intuitive than Excel's method, which requires cumulative probabilities and a lookup function.

As an example, suppose a manager estimates that the demand for a particular brand of television during the coming month will be 10, 15, 20, or 25, with respective probabilities 0.1, 0.3, 0.4, and 0.2. This typical discrete distribution is illustrated in Figure 15.14.

Generating Random Numbers with Excel Functions

The Discrete sheet of the **Probability Distributions.xlsx** file indicates how to work with a discrete distribution. (See Figure 15.15.) As you can see, there are two quite different ways to generate a random number from this distribution. We discuss the Excel way in detail in Section 15-4. For now, we simply mention that this is one case (of many) where it is much easier to generate random numbers with @RISK functions than with built-in Excel functions. Assuming that @RISK is loaded, all you need to do is enter the function RISKDISCRETE with two arguments, a list of possible values and a list of their probabilities, as in

=RISKDISCRETE(B11:B14,C11:C14)

The Excel way, which requires cumulative probabilities and a lookup table, takes more work and is harder to remember.

Figure 15.14 Discrete Distribution (from @RISK)

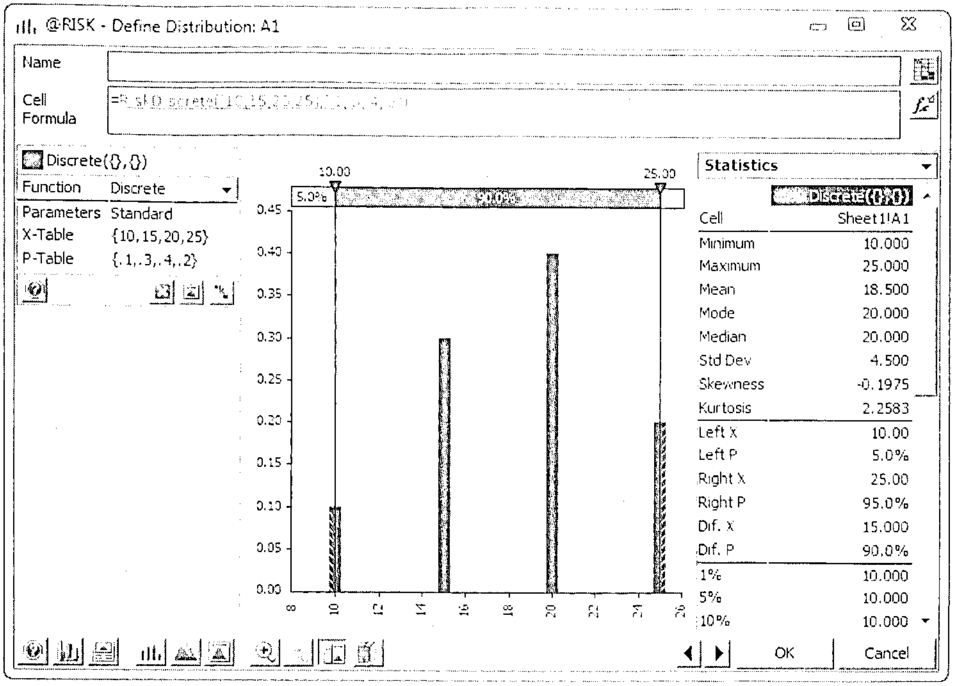


Figure 15.15 Properties of a Discrete Distribution

	A	B	C	D	E	F	G	H
1	**General discrete distribution**							
2								
3	**Characteristics**							
4	Discrete							
5	Can be symmetric or skewed (or bumpy, i.e., basically any shape)							
6	Bounded in both directions							
7	Not necessarily positive (depends on possible values)							
8								
9	**Parameters**				Lookup table required for Excel method			
10		Values	Probabilities		CumProb	Value		
11		10	0.1		0	10		
12		15	0.3		0.1	15		
13		20	0.4		0.4	20		
14		25	0.2		0.8	25		
15								
16	**Excel**		Example					
17	=VLOOKUP(RAND(),LookupTable,2)		25					
18								
19	**@RISK**							
20	=RISKDISCRETE(Values,Probs)		15					

This can have any shape, depending on the list of possible values and their probabilities.

@RISK Function: RISKDISCRETE

To generate a random number from any discrete probability distribution, enter the formula ***=RISKDISCRETE(valRange,probRange)*** *into any cell. Here valRange is the range with the possible values, and probRange is the range with their probabilities.*

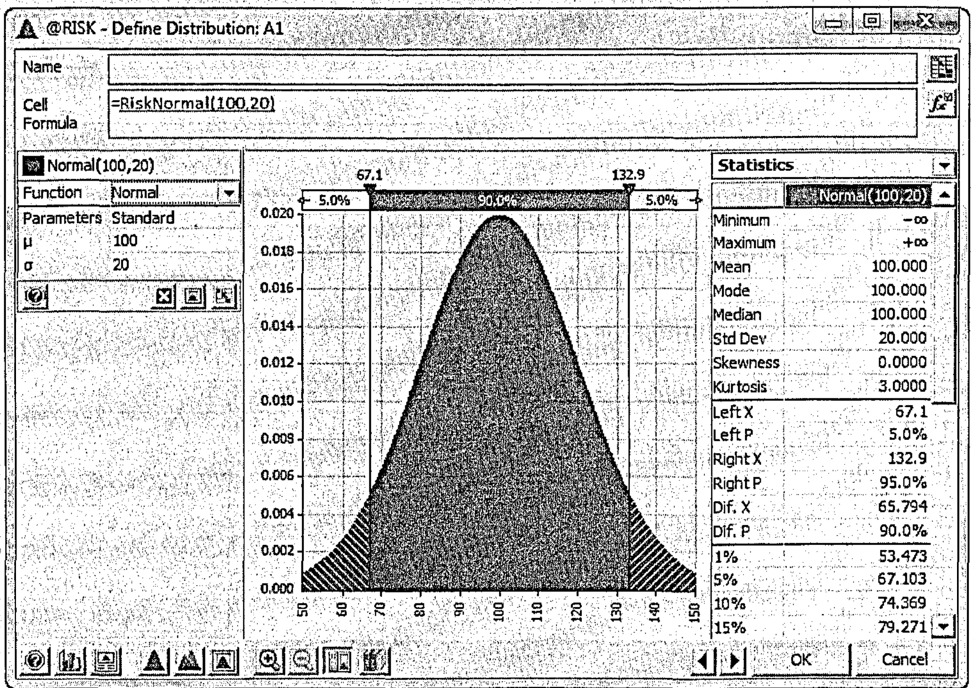
The selected input distributions for any simulation model reflect historical data and an analyst's best judgment as to what will happen in the future.

At this point, a relevant question is why a manager would choose this particular discrete distribution. First, it is clearly an approximation. After all, if it is possible to have demands of 20 and 25, it should also be possible to have demands between these values. Here, the manager approximates a discrete distribution with *many* possible values—all integers from 0 to 50, say—with a discrete distribution with a few typical values. This is fairly common in simulation modeling. Second, where do the probabilities come from? They are probably a blend of historical data (perhaps demand was near 15 in 30% of previous months) and the manager's subjective feelings about demand *next* month.

Normal Distribution

The *normal distribution* is the familiar bell-shaped curve that was discussed in detail in Chapter 5. (See Figure 15.16.) It is useful in simulation modeling as a continuous input distribution. However, it is *not* always the most appropriate distribution. It is symmetric, which can be a drawback when a skewed distribution is more realistic. Also, it allows negative values, which are not appropriate in many situations.

Figure 15.16
Normal Distribution (from @RISK)



A tip-off that a normal distribution might be an appropriate candidate for an input variable is a statement such as, "We believe the most likely value of demand is 100, and the chances are about 95% that demand will be no more than 40 units on either of side of this most likely value." Because a normally distributed value is within two standard deviations of its mean with probability 0.95, this statement translates easily to a mean of 100 and a standard deviation of 20. This does not imply that a normal distribution is the *only* candidate for the distribution of demand, but the statement naturally leads to this distribution.

The Normal sheet in the **Probability Distributions.xlsx** file indicates how you can generate normally distributed random numbers in Excel, either with or without @RISK. (See Figure 15.17.) This is one case where an add-in is not really necessary. The formula

=NORM.INV(RAND(),*Mean,Stdev*)

always works. Still, this is not as easy to remember as @RISK's formula

=RISKNORMAL,(*Mean,Stdev*)

@RISK Function: RISKNORMAL

To generate a normally distributed random number, enter the formula ***=RISKNORMAL (Mean,Stdev)*** *in any cell. Here, Mean and Stdev are the mean and standard deviation of the normal distribution.*

Figure 15.17 Properties of the Normal Distribution

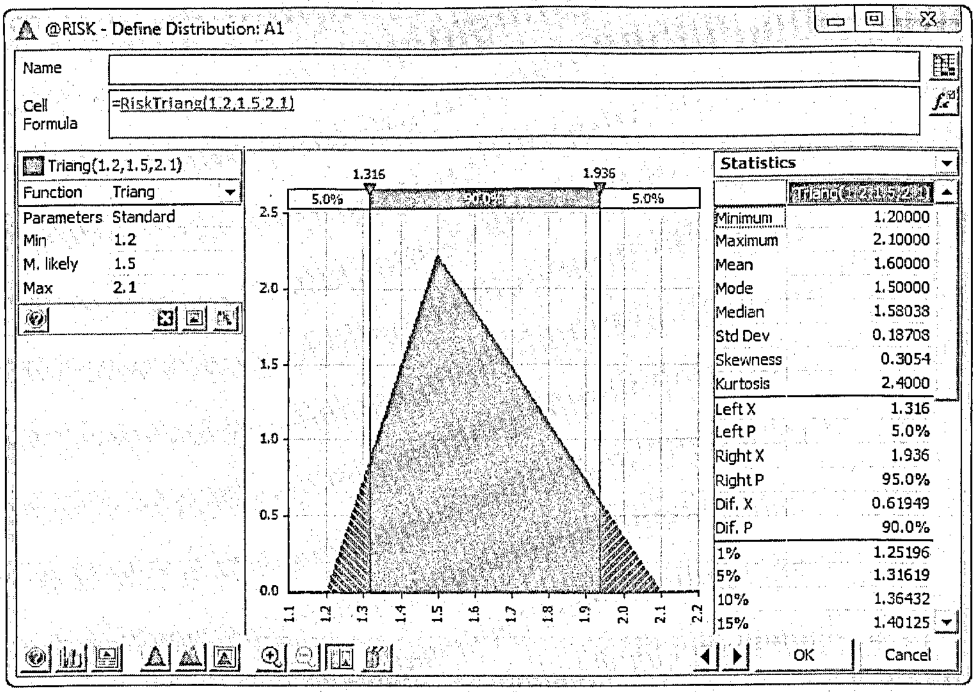
	A	B	C	D	E	F
1	**Normal distribution**					
2						
3	**Characteristics**					
4	Continuous			This is the familiar bell-shaped curve, defined by two parameters: the mean and the standard deviation.		
5	Symmetric (bell-shaped)					
6	Unbounded in both directions					
7	Is both positive and negative					
8						
9	**Parameters**					
10	Mean	100				
11	Stdev	10				
12						
13	**Excel**		Example			
14	=NORMINV(RAND(),Mean,Stdev)		83.60485164			
15						
16	**@RISK**					
17	=RISKNORMAL(Mean,Stdev)		85.03237755			

Triangular Distribution

The **triangular distribution** is somewhat similar to the normal distribution in that its density function rises to some point and then falls, but it is more flexible and intuitive than the normal distribution. Therefore, it is an excellent candidate for many continuous input variables. The shape of a triangular density function is literally a triangle, as shown in Figure 15.18. It is specified by three easy-to-understand parameters: the minimum possible value, the most likely value, and the maximum possible value. The high point of the triangle is above the most likely value. Therefore, if a manager states, "We believe the most likely development cost is \$1.5 million, and we don't believe the development cost could possibly be less than \$1.2 million or greater than \$2.1 million," the triangular distribution with these three parameters is a natural choice. As in this numerical example, note that the triangular distribution can be skewed if the mostly likely value is closer to one extreme than another. Of course, it can also be symmetric if the most likely value is right in the middle.

A triangular distribution is a good choice in many simulation models because it can have a variety of shapes and its parameters are easy to understand.

Figure 15.18
Triangular Distribution (from @RISK)



The Triangular sheet of the **Probability Distributions.xlsx** file indicates how to generate random values from this distribution. (See Figure 15.19.) As you can see, there is no easy way to do it with Excel functions only. However, it is easy with @RISK, using the RISKTRIANG function, as in

=RISKTRIANG(B10,B11,B12)

This function takes three arguments: the minimum value, the most likely value, and the maximum value—in this order and separated by commas. You will see this function in many of our examples. Just remember that it has an abbreviated spelling: RISKTRIANG, not RISKTRIANGULAR.

Figure 15.19
Properties of the Triangular Distribution

	A	B	C
1	**Triangular distribution**		
2			
3	**Characteristics**		
4	Continuous		
5	Can be symmetric or skewed in either direction		
6	Bounded in both directions		
7	Not necessarily positive (depends on bounds)		
8			
9	**Parameters**		
10	Min	50	
11	MostLikely	85	
12	Max	100	
13			
14	**Excel**		
15	There is no easy way to do it except by writing a macro.		
16			
17	**@RISK**		**Example**
18	=RISKTRIANG(Min,MostLikely,Max)		89.8069462

The density of this distribution is literally a triangle. The "top" of the triangle is above the most likely value, and the base of the triangle extends from the minimum value to the maximum value. It is intuitive for users because the three parameters have a natural meaning.

@RISK Function: RISKTRIANG

To generate a random number from a triangular distribution, enter the formula ***=RISKTRIANG(MinVal,MLVal,MaxVal)*** *in any cell. Here, MinVal is the minimum possible value, MLVal is the most likely value, and MaxVal is the maximum value.*

Binomial Distribution

The *binomial distribution* is a discrete distribution that was discussed extensively in Chapter 5. Recall that the binomial distribution applies to a very specific situation: when a number of independent and identical trials occur, and each trial results in a *success* or *failure*. Then the binomial random number is the number of successes in these trials. The two parameters of this distribution, n and p, are the number of trials and the probability of success on each trial.

A random number from a binomial distribution indicates the number of successes in a certain number of identical trials.

As an example, suppose an airline company sells 170 tickets for a flight and estimates that 80% of the people with tickets will actually show up for the flight. How many people will actually show up? It is tempting to state that *exactly* 80% of 170, or 136 people, will show up, but this neglects the inherent randomness. A more realistic way to model this situation is to say that each of the 170 people, independently of one another, will show up with probability 0.8. Then the number of people who actually show up is then binomially distributed with $n = 170$ and $p = 0.8$. (This assumes independent behavior across passengers, which might not be the case, for example, if whole families either show up or don't.) This distribution is illustrated in Figure 15.20.

The Binomial sheet of the **Probability Distributions.xlsx** file indicates how to generate random numbers from this distribution. (See Figure 15.21.) Although it is possible to do this with Excel using the built-in BINOM.INV function and the RAND function, it is not very intuitive or easy to remember. Clearly, the @RISK way is preferable. In the airline example, you would generate the number who show up with the formula

=RISKBINOMIAL(170,0.8)

Figure 15.20
Binomial Distribution (from @RISK)

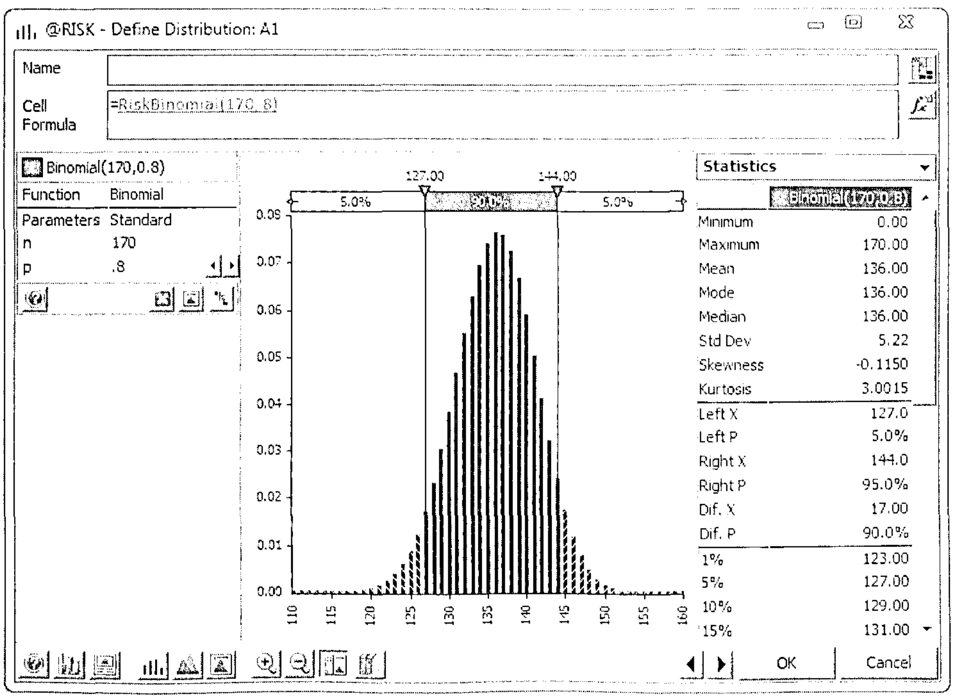


Figure 15.21 Properties of the Binomial Distribution

	A	B	C	D	E	F	G
1	**Binomial distribution**						
2							
3	**Characteristics**						
4	Discrete						
5	Can be symmetric or skewed						
6	Bounded below by 0, bounded above by Ntrials						
7	Nonnegative						
8							
9	**Parameters**						
10	NTrials	170					
11	PSuccess	0.8					
12							
13	**Excel**		**Example**				
14	=BINOM.INV(NTrials,PSuccess,RAND())		138				
15							
16	**@RISK**						
17	=RISKBINOMIAL(NTrials,PSuccess)		130				

This distribution is of the number of "successes" in a given number of identical, independent trials, when the probability of success is constant on each trial.

Note that the histogram in this figure is approximately bell-shaped. This is no accident. When the number of trials n is reasonably large and p isn't too close to 0 or 1, the binomial distribution can be well approximated by the normal distribution.

@RISK Function: RISKBINOMIAL

To generate a random number from a binomial distribution, enter the formula ***=RISKBINOMIAL(NTrials,PSuccess)*** *in any cell. Here, NTrials is the number of trials, and PSuccess is the probability of a success on each trial.*

It is natural to ask which distribution to use for a given uncertain quantity such as the price of oil, the demand for laptops, and so on. Admittedly, the choices we make in later examples are sometimes fairly arbitrary. However, in real business situations the choice is not always clear-cut, and it can make a difference in the results. Stanford professor Sam Savage and two of his colleages discuss this choice in a series of two articles on "Probability Management." (These articles are available online at http://lionhrtpub.com/orms/orms-2-06/frprobability.html and http://lionhrtpub.com/orms/orms-4-06/frprobability.html.) They argue that with the increasing importance of simulation models in today's business world, input distributions should not only be chosen carefully, but they should be kept and maintained as important corporate assets. They shouldn't just be chosen in some ad hoc fashion every time they are needed. For example, if the price of oil is an important input in many of a company's decisions, then experts within the company should assess an appropriate distribution for the price of oil and modify it as necessary when new information arises. The authors even suggest a new company position, Chief Probability Officer, to control access to the company's probability distributions.

So as you are reading these final two chapters, keep Savage's ideas in mind. The choice of probability distributions for inputs is not easy, but it is also not arbitrary. The choice *can* make a difference in the results. This is the reason why you want as many families of probability distributions in your toolbox as possible. You then have more flexibility in choosing a distribution that is appropriate for your situation.

PROBLEMS

Note: Student solutions for problems whose numbers appear within a colored box are available for purchase at www.cengagebrain.com.

Level A

1. Use the RAND function and the Copy command to generate a set of 100 random numbers.
   a. What fraction of the random numbers are smaller than 0.5?
   b. What fraction of the time is a random number less than 0.5 followed by a random number greater than 0.5?
   c. What fraction of the random numbers are larger than 0.8?
   d. Freeze these random numbers. However, instead of pasting them over the original random numbers, paste them onto a new range. Then press the F9 recalculate key. The original random numbers should change, but the pasted copy should remain the same.
2. Use Excel's functions (not @RISK) to generate 1000 random numbers from a normal distribution with mean 100 and standard deviation 10. Then freeze these random numbers.
   a. Calculate the mean and standard deviation of these random numbers. Are they approximately what you would expect?
   b. What fraction of these random numbers are within k standard deviations of the mean? Answer for $k = 1$; for $k = 2$; for $k = 3$. Are the answers close to what they should be (according to the empirical rules you learned in Chapters 2 and 5)?
   c. Create a histogram of the random numbers using 10 to 15 categories of your choice. Does this histogram have approximately the shape you would expect?
3. Use @RISK's Define Distributions tool to show a uniform distribution from 400 to 750. Then answer the following questions.
   a. What are the mean and standard deviation of this distribution?
   b. What are the 5th and 95th percentiles of this distribution?
   c. What is the probability that a random number from this distribution is less than 450?
   d. What is the probability that a random number from this distribution is greater than 650?
   e. What is the probability that a random number from this distribution is between 500 and 700?
4. Use @RISK's Define Distributions tool to draw a normal distribution with mean 500 and standard deviation 100. Then answer the following questions.
   a. What is the probability that a random number from this distribution is less than 450?
   b. What is the probability that a random number from this distribution is greater than 650?
   c. What is the probability that a random number from this distribution is between 500 and 700?
5. Use @RISK's Define Distributions tool to show a triangular distribution with parameters 300, 500, and 900. Then answer the following questions.
   a. What are the mean and standard deviation of this distribution?
   b. What are the 5th and 95th percentiles of this distribution?
   c. What is the probability that a random number from this distribution is less than 450?
   d. What is the probability that a random number from this distribution is greater than 650?
   e. What is the probability that a random number from this distribution is between 500 and 700?
6. Use @RISK's Define Distributions tool to show a binomial distribution that results from 50 trials with probability of success 0.3 on each trial, and use it to answer the following questions.
   a. What are the mean and standard deviation of this distribution?
   b. You have to be more careful in interpreting @RISK probabilities with a discrete distribution such as this binomial. For example, if you move the left slider to 11, you find a probability of 0.139 to the left of it. But is this the probability of "less than 11" or "less than or equal to 11"? One way to check is to use Excel's BINOMDIST function. Use this function to interpret the 0.139 value from @RISK.
   c. Using part **b** to guide you, use @RISK to find the probability that a random number from this

distribution will be greater than 17. Check your answer by using the BINOMDIST function appropriately in Excel.

7. Use @RISK's Define Distributions tool to draw a triangular distribution with parameters 200, 300, and 600. Then superimpose a normal distribution on this drawing, choosing the mean and standard deviation to match those from the triangular distribution. (Click the Add Overlay button at the bottom of the window and then choose the distribution to superimpose.)
   a. What are the 5th and 95th percentiles for these two distributions?
   b. What is the probability that a random number from the triangular distribution is less than 400? What is this probability for the normal distribution?
   c. Experiment with the sliders to answer questions similar to those in part **b**. Would you conclude that these two distributions differ most in the extremes (right or left) or in the middle? Explain.

8. We all hate to keep track of small change. By using random numbers, it is possible to eliminate the need for change and give the store and the customer a fair deal. This problem indicates how it could be done.
   a. Suppose that you buy something for $0.20. How could you use random numbers (built into the cash register system) to decide whether you should pay $1.00 or nothing?
   b. If you bought something for $9.60, how would you use random numbers to eliminate the need for change?
   c. In the long run, why is this method fair to both the store and the customers? Would you personally (as a customer) be willing to abide by such a system?

Level B

9. A company is about to develop and then market a new product. It wants to build a simulation model for the entire process, and one key uncertain input is the development cost. For each of the following scenarios, choose an appropriate distribution together with its parameters, justify your choice in words, and use @RISK's Define Distributions tool to show your chosen distribution.
   a. Company experts have no idea what the distribution of the development cost is. All they can state is "we are 95% sure it will be at least $450,000, and we are 95% sure it will be no more than $650,000."
   b. Company experts can still make the same statement as in part **a**, but now they can also state: "We believe the distribution is symmetric, reasonably bell-shaped, and its most likely value is about $550,000."
   c. Company experts can still make the same statement as in part **a**, but now they can also state: "We believe the distribution is skewed to the right, and its most likely value is about $500,000."

10. Continuing the preceding problem, suppose that another key uncertain input is the development time, which is measured in an *integer* number of months. For each of the following scenarios, choose an appropriate distribution together with its parameters, justify your choice in words, and use @RISK's Define Distributions tool to show your chosen distribution.
   a. Company experts believe the development time will be from 6 to 10 months, but they have absolutely no idea which of these will result.
   b. Company experts believe the development time will be from 6 to 10 months. They believe the probabilities of these five possible values will increase linearly to a most likely value at 8 months and will then decrease linearly.
   c. Company experts believe the development time will be from 6 to 10 months. They believe that 8 months is twice as likely as either 7 months or 9 months and that either of these latter possibilities is three times as likely as either 6 months or 10 months.

15-3 SIMULATION AND THE FLAW OF AVERAGES

To help motivate simulation modeling in general, we present a simple example in this section. It will clearly show the distinction between Figure 15.1 (a deterministic model with best-guess inputs) and Figure 15.2 (an appropriate simulation model). In doing so, it will illustrate a pitfall called the "flaw of averages."[6]

[6] As far as we know, the term "flaw of averages" was coined by Sam Savage, the same Stanford professor quoted earlier.

EXAMPLE 15.1 ORDERING CALENDARS AT WALTON BOOKSTORE

The Flaw of Averages

In August, Walton Bookstore must decide how many of next year's nature calendars to order. Each calendar costs the bookstore \$7.50 and sells for \$10. After January 1, all unsold calendars will be returned to the publisher for a refund of \$2.50 per calendar. Walton believes that the number of calendars it can sell by January 1 follows some probability distribution with mean 200. Walton believes that ordering to the average demand, that is, ordering 200 calendars, is a good decision. Is it?

Objective To illustrate the difference between a deterministic model with a best guess for uncertain inputs and a simulation model that incorporates uncertainty explicitly.

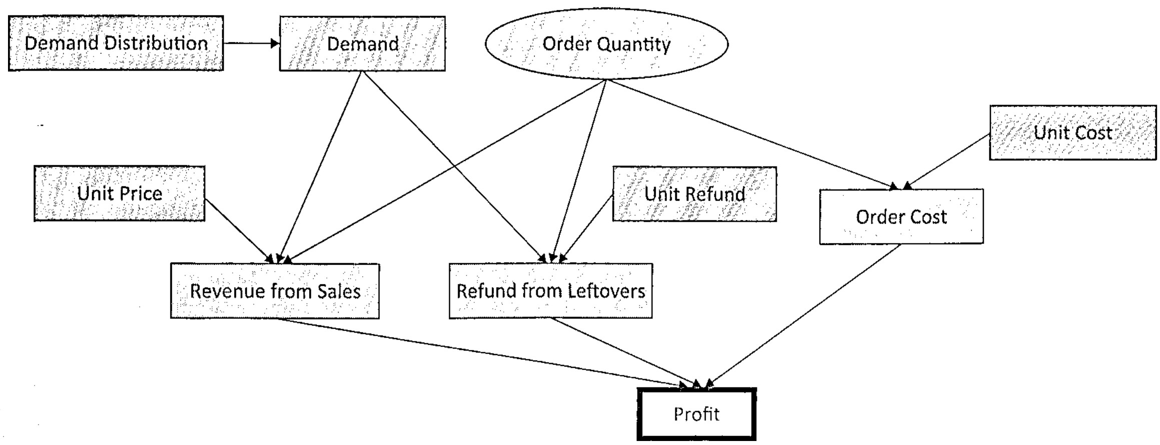
Where Do the Numbers Come From?

The monetary values are straightforward. The mean demand is probably an estimate based on historical demands for similar calendars.

Solution

The variables for this model are shown in Figure 15.22. (See the file **Ordering Calendars Big Picture.xlsx**.) Note that in addition to the "Big Picture" conventions illustrated in the two previous chapters, we use a green rectangle with a rounded top for uncertain quantities. (For most of the rest of this chapter, you can refer back to this diagram.) An order quantity is chosen and demand is then observed. The ordering cost is based on the order quantity, the revenue is based on the smaller of the order quantity and demand, and there is a refund if demand is less than the order quantity.

Figure 15.22 Big Picture for Ordering Model



A deterministic model appears in Figure 15.23. (See the file **Ordering Calendars - Flaw of Averages.xlsx**.) Assuming the best guess for demand, Walton orders to this average value, and it appears that the company's best guess for profit is \$500. (The formulas in cells B16:F16 are straightforward and are listed in row 18. Before reading further, do you believe that the *average* profit will be \$500 when uncertainty in demand is introduced explicitly (and the company still orders 200 calendars)? Think what happens to profit when demand is less than 200 and when it is greater than 200. Are these two cases symmetric?

Figure 15.23 Deterministic Model

	A	B	C	D	E	F
1	**Walton's bookstore - deterministic model**					
2						
3	**Cost data**					
4	Unit cost	$7.50				
5	Unit price	$10.00				
6	Unit refund	$2.50				
7						
8	**Uncertain quantity**					
9	Demand (average shown)	200				
10						
11	**Decision variable**					
12	Order quantity	200				
13						
14	**Profit model**					
15		Demand	Revenue	Cost	Refund	Profit
16		200	$2,000.00	$1,500.00	$0.00	$500.00
17						
18	Formulas	=B11	=B7*MIN(B11,B14)	=B6*B14	=B8*MAX(B14-B11,0)	=C18-D18+E18

This deterministic model gives no hint of what will happen when demand is treated explicitly as random. From the simulation model on the next sheet, we see that the average profit is nowhere near the $500 value in this sheet.

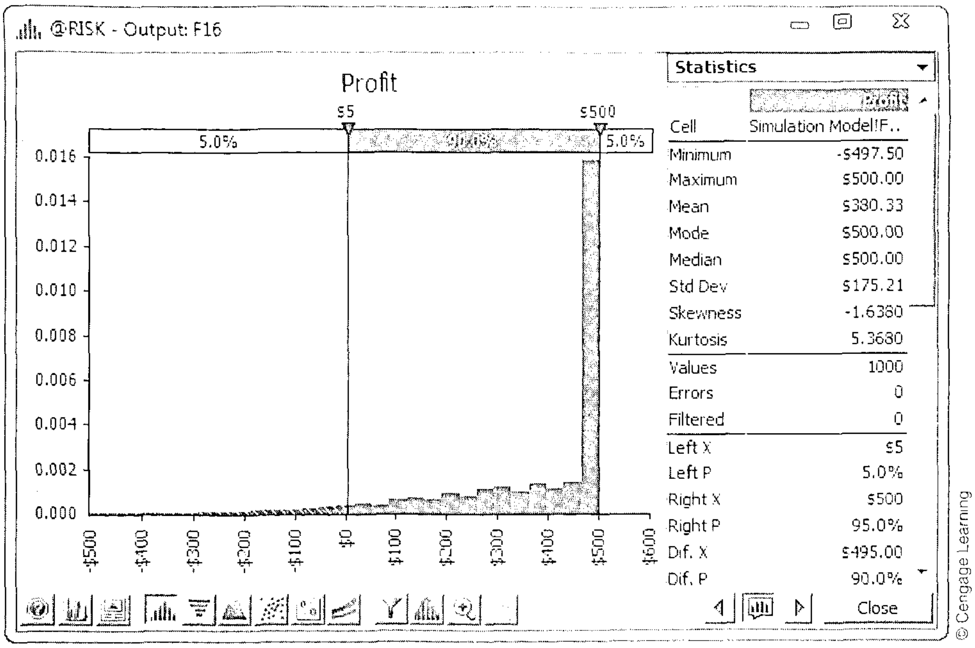
We now contrast this with a simulation model where the demand in cell B9 is replaced by a random number. For this example, we assume that demand is *normally* distributed with mean 200 and standard deviation 40, although these specific assumptions are not crucial for the qualitative aspects of the example. Specifically, cell B9 should contain the formula **=ROUND(RISKNORMAL(200,40),0)** where the ROUND function has been used to round to the nearest integer. (We assume that @RISK has been loaded.) Now the model appears as in Figure 15.24.

The random demand in cell B9 is now live, as are its dependents in row 16, so each time you press the F9 key, you get a new demand and associated profit. (This assumes that the @RISK "dice" button is toggled to colored, its "random" setting. More will be said about this setting later in the chapter.) Do you get about $500 in profit on average? Absolutely not! The situation isn't symmetric. The *largest* profit you can get is $500, which occurs about half the time, whenever demand is greater than 200. A typical such situation appears in the figure, where the excess demand of 54 is simply lost. However, when demand is less than 200, the profit is *less than* $500, and it keeps decreasing as demand decreases.

Figure 15.24 Simulation Model

	A	B	C	D	E	F
1	**Walton's bookstore - simulation model**					
2						
3	**Cost data**					
4	Unit cost	$7.50				
5	Unit price	$10.00				
6	Unit refund	$2.50				
7						
8	**Uncertain quantity (assumed normal with mean 200, stdev 40)**					
9	Demand (random)	254				
10						
11	**Decision variable**					
12	Order quantity	200				
13						
14	**Profit model**					
15		Demand	Revenue	Cost	Refund	Profit
16		254	$2,000.00	$1,500.00	$0.00	$500.00

Figure 15.25
Histogram of Simulated Profits



We ran @RISK with 1000 iterations (which will be explained in detail in Section 15-5) and found the resulting histogram of 1000 simulated profits shown in Figure 15.25. The large spike on the right is due to the cases where demand is 200 or more and profit is \$500. All the little spikes to the left are where demand is less than 200 and profit is less than \$500, sometimes considerably less. You can see on the right that the *mean* profit, the average of the 1000 simulated profits, is only about \$380, well less than the \$500 suggested by the deterministic model.

The point of this simple example is that a deterministic model can be very misleading. In particular, the output from a deterministic model that uses best guesses for uncertain inputs is *not* necessarily equal to, or even close to, the average of the outputs from a simulation. This is exactly what "the flaw of averages" means. ▪

FUNDAMENTAL INSIGHT

The Flaw of Averages

If a model contains uncertain inputs, it can be very misleading to build a deterministic model by using the *means* of the inputs to predict an output. The resulting output value can be considerably different—lower *or* higher—than the mean of the output values obtained from running a simulation with uncertainty incorporated explicitly.

15-4 SIMULATION WITH BUILT-IN EXCEL TOOLS

In this section, we show how spreadsheet simulation models can be developed and analyzed with Excel's built-in tools without using add-ins. As you will see, this is certainly possible, but it presents two problems. First, the @RISK functions illustrated in the **Probability Distributions.xlsx** file are not available. You are able to use only Excel's RAND function and transformations of it to generate random numbers from various probability distributions. (You can also use the RANDBETWEEN function, but except for special cases, this doesn't help much.) Second, there is a bookkeeping problem. Once you build an Excel model with output cells linked to appropriate random input cells, you can press the F9 key as often as you like to see how the outputs vary. However, there is no quick way to keep track of these output values and summarize them. This bookkeeping feature is the real strength of a simulation add-in such as @RISK. It can be done with Excel, usually with data tables, but the summarization of the resulting data is completely up to the user—you. Therefore, we strongly recommend that you use the "Excel-only" method described in this section only if you don't have an add-in such as @RISK.

To illustrate the Excel-only procedure, we continue to analyze the calendar problem from Example 15.1. This general problem occurs when a company (such as a news vendor) must make a one-time purchase of a product (such as a newspaper) to meet customer demands for a certain period of time. If the company orders too few newspapers, it will lose potential profit by not having enough on hand to satisfy its customers. If it orders too many, it will have newspapers left over at the end of the day that, at best, can be sold at a loss. More generally, the problem is to match supply to an uncertain demand, a very common problem in business. In much of the rest of this chapter, we will discuss variations of this problem, generally referred to as the *newsvendor* problem.

EXAMPLE 15.2 SIMULATING WITH EXCEL ONLY AT WALTON BOOKSTORE

Recall that Walton Bookstore must decide how many of next year's nature calendars to order. Each calendar costs the bookstore $7.50 and sells for $10. After January 1, all unsold calendars will be returned to the publisher for a refund of $2.50 per calendar. In this version, we assume that demand for calendars (at the full price) is given by the probability distribution shown in Table 15.1. Walton wants to develop a simulation model to help it decide how many calendars to order.

Table 15.1 Probability Distribution of Demand for Walton Example

Demand	Probability
100	0.30
150	0.20
200	0.30
250	0.15
300	0.05

Objective To use built-in Excel tools—including the RAND function and data tables, but no add-ins—to simulate profit for several order quantities and ultimately choose the "best" order quantity.

Where Do the Numbers Come From?

The numbers in Table 15.1 are the key to the simulation model. They are discussed in more detail next.

Solution

We first discuss the probability distribution in Table 15.1. It is a discrete distribution with only five possible values: 100, 150, 200, 250, and 300. In reality, it is clear that other values of demand are possible. For example, there could be demand for exactly 187 calendars. In spite of its apparent lack of realism, we use this discrete distribution for two reasons. First, its simplicity is a nice feature to get you started with simulation modeling. Second, discrete distributions are often used in real business simulation models. Even though the discrete distribution is only an *approximation* to reality, it can still provide important insights into the actual problem.

As for the probabilities listed in Table 15.1, they are typically drawn from historical data or (if historical data are lacking) educated guesses. In this case, the manager of Walton Bookstore has presumably looked at demands for calendars in previous years, and he has used any information he has about the market for next year's calendars to estimate, for example, that the probability of a demand for 200 calendars is 0.30. The five probabilities in this table *must* sum to 1. Beyond this requirement, they should be as reasonable and consistent with reality as possible.

It is important to realize that this is really a decision problem under uncertainty. Walton must choose an order quantity *before* knowing the demand for calendars. Unfortunately, Solver cannot be used because of the uncertainty.[7] Therefore, we develop a simulation model for any *fixed* order quantity. Then we run this simulation model with various order quantities to see which one appears to be best.

Developing the Simulation Model

Developing the Walton Model with Excel Tools Only

Now we discuss the ordering model. For any fixed order quantity, we show how Excel can be used to simulate 1000 replications (or any other number of replications). Each replication is an independent replay of the events that occur. To illustrate, suppose you want to simulate profit if Walton orders 200 calendars. Figure 15.26 illustrates the results obtained

Figure 15.26 Walton Bookstore Simulation Model

	A	B	C	D	E	F	G	H	I	J	K
1	**Simulation of Walton's bookstore**							**Range names used:**			
2								LookupTable	=Model!$D$5:$F$9		
3	**Cost data**			**Demand distribution**				Order_quantity	=Model!$B$9		
4	Unit cost	$7.50		Cum Prob	Demand	Probability		Profit	=Model!$G$19:$G$1018		
5	Unit price	$10.00		0.00	100	0.30		Unit_cost	=Model!$B$4		
6	Unit refund	$2.50		0.30	150	0.20		Unit_price	=Model!$B$5		
7				0.50	200	0.30		Unit_refund	=Model!$B$6		
8	**Decision variable**			0.80	250	0.15					
9	Order quantity	200		0.95	300	0.05					
10											
11	**Summary measures for simulation below**										
12	Average profit	$204.13									
13	Stdev of profit	$328.04									
14	Minimum profit	−$250.00									
15	Maximum profit	$500.00									
16											
17	**Simulation**								**Distribution of profit**		
18	Replication	Random #	Demand	Revenue	Cost	Refund	Profit		Value	Frequency	Rel. Freq.
19	1	0.4695	150	$1,500	$1,500	$125	$125		−250	299	0.299
20	2	0.0022	100	$1,000	$1,500	$250	−$250		125	191	0.191
21	3	0.2614	100	$1,000	$1,500	$250	−$250		500	510	0.51
22	4	0.6220	200	$2,000	$1,500	$0	$500				
23	5	0.1417	100	$1,000	$1,500	$250	−$250				
1016	998	0.1005	100	$1,000	$1,500	$250	−$250				
1017	999	0.3798	150	$1,500	$1,500	$125	$125				
1018	1000	0.4530	150	$1,500	$1,500	$125	$125				
1019											

[7] @RISK contains a tool called RISKOptimizer that can be used for optimization in a simulation model, but we will not discuss it here.

by simulating 1000 independent replications for this order quantity. (See the file **Ordering Calendars - Excel Only 1.xlsx.**) Note that there are many hidden rows in Figure 15.26. To develop this model, use the following steps.

1. **Inputs.** Enter the cost data in the range B4:B6, the probability distribution of demand in the range E5:F9, and the proposed order quantity, 200, in cell B9. Pay particular attention to the way the probability distribution is entered (and compare to the Discrete sheet in the **Probability Distributions.xlsx** file). Columns E and F contain the possible demand values and the probabilities from Table 15.1. It is also necessary (see step 2 for the reasoning) to have the cumulative probabilities in column D. To obtain these, first enter the value 0 in cell D5. Then enter the formula

   =F5+D5

   in cell D6 and copy it to the range D7:D9.

2. **Generate random demands.** The key to the simulation is the generation of a customer demand in column C from a random number generated by the RAND function in column B and the probability distribution of demand. Here is how it works. The interval from 0 to 1 is split into five segments: 0.0 to 0.3 (length 0.3), 0.3 to 0.5 (length 0.2), 0.5 to 0.8 (length 0.3), 0.8 to 0.95 (length 0.15), and 0.95 to 1.0 (length 0.05). Note that these lengths are the probabilities of the various demands. Then a demand is associated with each random number, depending on which interval the random number falls in. For example, if a random number is 0.5279, this falls in the third interval, so it is associated with the third possible demand value, 200.

   This rather cumbersome procedure for generating a discrete random number is not necessary when you use @RISK.

   To implement this procedure, you use a VLOOKUP function based on the range D5:F9 (named LookupTable). This table has the cumulative probabilities in column D and the possible demand values in column E. In fact, the whole purpose of the cumulative probabilities in column D is to allow the use of the VLOOKUP function. To generate the simulated demands, enter the formula

   =VLOOKUP(RAND(),LookupTable,2)

   in cell C19. This formula compares any RAND value to the values in D5:D9 and returns the appropriate demand from E5:E9. (In the file, you will note that random cells are colored green. This coloring convention is not required, but we use it consistently to identify the random cells.)

   This step is the key to the simulation, so make sure you understand exactly what it entails. The rest is bookkeeping, as indicated in the following steps.

3. **Revenue.** Once the demand is known, the number of calendars sold is the smaller of the demand and the order quantity. For example, if 150 calendars are demanded, 150 will be sold. But if 250 are demanded, only 200 can be sold (because Walton orders only 200). Therefore, to calculate the revenue in cell D19, enter the formula

   =Unit_price*MIN(C19,Order_quantity)

4. **Ordering cost.** The cost of ordering the calendars does not depend on the demand; it is the unit cost multiplied by the number ordered. Calculate this cost in cell E19 with the formula

   =Unit_cost*Order_quantity

5. **Refund.** If the order quantity is greater than the demand, there is a refund of $2.50 for each calendar left over; otherwise, there is no refund. Therefore, calculate the refund in cell F19 with the formula

   =Unit_refund*MAX(Order_quantity-C19,0)

For example, if demand is 150, then 50 calendars are left over, and this MAX is 50, the larger of 50 and 0. However, if demand is 250, then no calendars are left over, and this MAX is 0, the larger of −50 and 0. (This calculation could also be accomplished with an IF function instead of a MAX function.)

6. **Profit.** Calculate the profit in cell G19 with the formula

   =D19+F19-E19

7. **Copy to other rows.** This is a "one-line" simulation, where all of the logic is captured in a single row, row 19. For one-line simulations, you can replicate the logic with new random numbers very easily by copying down. Copy row 19 down to row 1018 to generate 1000 replications.

8. **Summary measures.** Each profit value in column G corresponds to one randomly generated demand. You usually want to see how these vary from one replication to another. First, calculate the average and standard deviation of the 1000 profits in cells B12 and B13 with the formulas

   =AVERAGE(G19:G1018)

   and

   =STDEV.S(G19:G1018)

   Similarly, calculate the smallest and largest of the 1000 profits in cells B14 and B15 with the MIN and MAX functions.

9. **Distribution of simulated profits.** There are only three possible profits, −$250, $125, or $500 (depending on whether demand is 100, 150, or at least 200—see the following discussion). You can use the COUNTIF function to count the number of times each of these possible profits is obtained. To do so, enter the formula

   =COUNTIF($G$19:$G$1018,I19)

   in cell J19 and copy it down to cell J21.

Checking Logic with Deterministic Inputs

It can be difficult to check whether the logic in your model is correct, because of the random numbers. The reason is that you usually get different output values, depending on the particular random numbers generated. Therefore, it is sometimes useful to enter well-chosen *fixed* values for the random inputs, just to see whether your logic is correct. We call these *deterministic checks*. In the present example, you might try several fixed demands, at least one of which is *less than* the order quantity and at least one of which is *greater than* the order quantity. For example, if you enter a fixed demand of 150, the revenue, cost, refund, and profit should be $1500, $1500, $125, and $125, respectively. Or if you enter a fixed demand of 250, these outputs are $2000, $1500, $0, and $500. There is no randomness in these values; every correct model should get these same values. If your model doesn't get these values, there must be a logic error in your model that has nothing to do with random numbers or simulation. Of course, you should fix any such logical errors before reentering the *random* demand and running the simulation.

You can make a similar check by keeping the random demand, repeatedly pressing the F9 key, and watching the outputs for the different random demands. For example, if the refund is not $0 every time demand exceeds the order quantity, you know you have a logical error in at least one formula. The advantage of deterministic checks is that you can compare your results with those of other users, using *agreed-upon test values* of the random quantities. You should all get exactly the same outputs.

Discussion of the Simulation Results

At this point, it is a good idea to stand back and see what you have accomplished. First, in the body of the simulation, rows 19 through 1018, you randomly generated 1000 possible demands and the corresponding profits. Because there are only five possible demand values (100, 150, 200, 250, and 300), there are only five possible profit values: –\$250, \$125, \$500, \$500, and \$500. Also, note that for the order quantity 200, the profit is \$500 regardless of whether demand is 200, 250, or 300. (Make sure you understand why.) A tally of the profit values in these rows, including the hidden rows, indicates that there are 299 rows with profit equal to –\$250 (demand 100), 191 rows with profit equal to \$125 (demand 150), and 510 rows with profit equal to \$500 (demand 200, 250, or 300). The average of these 1000 profits is \$204.13, and their standard deviation is \$328.04. (Again, however, remember that your answers will probably differ from these because of different random numbers.)

For this particular model, the output distribution is also discrete: There are only three possible profits for an order quantity of 200.

Typically, a simulation model should capture one or more output variables, such as profit. These output variables depend on random inputs, such as demand. The goal is to estimate the probability distributions of the outputs. In the Walton simulation the estimated probability distribution of profit is

$$P(\text{Profit} = -\$250) = 299/1000 = 0.299$$

$$P(\text{Profit} = \$125) = 191/1000 = 0.191$$

$$P(\text{Profit} = \$500) = 510/1000 = 0.510$$

The estimated mean of this distribution is \$204.13 and the estimated standard deviation is \$328.04. It is important to realize that if the entire simulation is run again with *different* random numbers (such as the ones you might have generated on your PC), the answers will probably be slightly different. For illustration, we pressed the F9 key five times and got the following average profits: \$213.88, \$206.00, \$212.75, \$219.50, and \$189.50. So this is truly a case of "answers will vary."

Notes about Confidence Intervals

The confidence interval provides a measure of accuracy of the mean profit, as estimated from the simulation.

It is common in computer simulations to estimate the mean of some distribution by the average of the simulated observations. The usual practice is then to accompany this estimate with a **confidence interval,** which indicates the accuracy of the estimate. You should recall from Chapter 8 that to obtain a confidence interval for the mean, you start with the estimated mean and then add and subtract a multiple of the *standard error* of the estimated mean. If the estimated mean (that is, the average) is $\overline{X}$, the confidence interval is given in the following formula.

Confidence Interval for the Mean

$$\overline{X} \pm \text{Multiple} \times \text{Standard Error of } \overline{X}$$

We repeat these basic facts about confidence intervals from Chapter 8 here for your convenience.

The standard error of $\overline{X}$ is the standard deviation of the observations divided by the square root of n, the number of observations:

Standard Error of $\bar{X}$

$$s/\sqrt{n}$$

Here, s is the symbol for the standard deviation of the observations. You can obtain it with the STDEV.S function in Excel.

The *multiple* in the confidence interval formula depends on the confidence level and the number of observations. If the confidence level is 95%, for example, the multiple is very close to 2, so a good guideline is to go out two standard errors on either side of the average to obtain an approximate 95% confidence interval for the mean.

The idea is to choose the number of iterations large enough so that the resulting confidence interval will be sufficiently narrow.

Approximate 95% Confidence Interval for the Mean

$$\bar{X} \pm 2s/\sqrt{n}$$

Analysts often plan a simulation so that the confidence interval for the mean of some important output will be sufficiently narrow. The reasoning is that narrow confidence intervals imply more precision about the estimated mean of the output variable. If the confidence level is fixed at some value such as 95%, the only way to narrow the confidence interval is to simulate more replications. Assuming that the confidence level is 95%, the following value of n is required to ensure that the resulting confidence interval will have a half-length approximately equal to some specified value B:

Sample Size Determination

$$n = \frac{4 \times (\text{Estimated standard deviation})^2}{B^2}$$

This formula requires an estimate of the standard deviation of the output variable. For example, in the Walton simulation the 95% confidence interval with $n = 1000$ has half-length ($224.46 − $183.79)/2 = $20.33. Suppose that you want to reduce this half-length to $12.50—that is, you want $B = \$12.50$. You do not know the exact standard deviation of the profit distribution, but you can estimate it from the simulation as $328.04. Therefore, to obtain the required confidence interval half-length B, you need to simulate n replications, where

$$n = \frac{4(328.04)^2}{12.50^2} \approx 2755$$

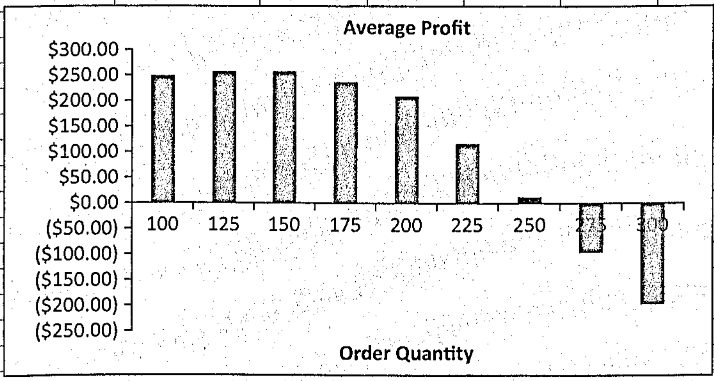
(When this formula produces a noninteger, it is common to round upward.) The claim, then, is that if you rerun the simulation with 2755 replications rather than 1000 replications, the half-length of the 95% confidence interval for the mean profit will be close to $12.50.

Finding the Best Order Quantity

We are not yet finished with the Walton example. So far, the simulation has been run for only a single order quantity, 200. Walton's ultimate goal is to find the *best* order quantity. Even this statement must be clarified. What does "best" mean? As in Chapter 6, one possibility is to use the *expected* profit—that is, EMV—as the optimality criterion, but other characteristics of the profit distribution could influence the decision. You can obtain the required outputs with a data table. Specifically, you can use a data table to rerun the simulation for other order quantities. This data table and a corresponding chart are shown in Figure 15.27. (This is still part of the finished version of the **Ordering Calendars - Excel Only 1.xlsx** file.)

Figure 15.27
Data Table for Walton Bookstore Simulation

	M	N	O	P	Q	R	S	T
17	Data table for average profit versus order quantity							
18	Order quantity	Average Profit						
19		$204.13						
20	100	$250.00						
21	125	$256.81						
22	150	$258.00						
23	175	$237.44						
24	200	$209.75						
25	225	$118.50						
26	250	$12.63						
27	275	($95.88)						
28	300	($196.13)						



To optimize in simulation models, try various values of the decision variable(s) and run the simulation for each of them.

To create this table, enter the trial order quantities shown in the range M20:M28, enter the link =**B12** to the average profit in cell N19, and select the data table range M19:N28. Then select Data Table from the What-If Analysis dropdown list on the Data ribbon, specifying that the column input cell is B9. (See Figure 15.26.) Finally, construct a column chart of the average profits in the data table. Note that an order quantity of 150 appears to maximize the average profit. Its average profit of $258.00 is slightly higher than the average profits from nearby order quantities and much higher than the profit gained from an order of 200 or more calendars. However, again keep in mind that this is a simulation, so that all of these average profits depend on the particular random numbers generated. If you rerun the simulation with different random numbers, it is conceivable that some other order quantity could be best.

Excel Tip: Calculation Settings with Data Tables

Sometimes you will create a data table and the values will be constant the whole way down. This could mean you did something wrong, but more likely it is due to a calculation setting. To check, go to the Formulas ribbon and click the Calculation Options dropdown arrow. If it isn't Automatic (the default setting), you need to click the Calculate Now (or Calculate Sheet) button or press the F9 key to make the data table calculate correctly. (The Calculate Now and F9 key recalculate everything in your workbook. The Calculate Sheet option recalculates only the active sheet.) Note that the Automatic Except for Data Tables setting is there for a reason.

Data tables, especially those based on complex simulations, can take a lot of time to recalculate, and with the default setting, this recalculation occurs every time anything changes in your workbook. So the Automatic Except for Data Tables setting is handy to prevent data tables from recalculating until you force them to by pressing the F9 key or clicking one of the Calculate buttons.

Using a Data Table to Repeat Simulations

The Walton simulation is a particularly simple one-line simulation model. All of the logic—generating a demand and calculating the corresponding profit—can be captured in a single row. Then to replicate the simulation, you can simply copy this row down as far as you like. Many simulation models are significantly more complex and require more than one row to capture the logic. Nevertheless, they still result in one or more output quantities (such as profit) that you want to replicate. We now illustrate another method of **replicating with Excel tools only** that is more general (still using the Walton example). It uses a data table to generate the replications. Refer to Figure 15.28 and the file **Ordering Calendars - Excel Only 2.xlsx**.

Figure 15.28
Using a Data Table to Simulate Replications

	A	B	C	D	E	F
17	**Simulation**					
18		Demand	Revenue	Cost	Refund	Profit
19		200	$2,000	$1,500	$0	$500
20						
21	**Data table for replications, each shows profit from that replication**					
22	Replication	Profit				
23		$500				
24	1	$125				
25	2	–$250				
26	3	$500				
27	4	–$250				
28	5	$125				
1021	998	$125				
1022	999	$500				
1023	1000	$500				

Through row 19, the only difference between this model and the previous model is that the RAND function is embedded in the VLOOKUP function for demand in cell B19. This makes the model slightly more compact. As before, it uses the given data at the top of the spreadsheet to construct a typical "prototype" of the simulation in row 19. This time, however, you do not copy row 19 down. Instead, you create a data table in the range A23:B1023 to replicate the basic simulation 1000 times. In column A, you list the replication numbers, 1 to 1000. Next, you enter the formula **=F19** in cell B23. This forms a link to the profit from the prototype row for use in the data table. Then you create a data table and enter *any blank cell* (such as C23) as the column input cell. (No row input cell is necessary, so its box should be left empty.) This tricks Excel into repeating the row 19 calculations 1000 times, each time with a new random number, and reporting the profits in column B of the data table. (If you wanted to see other simulated quantities, such as revenue, for each replication, you could add extra output columns to the data table.)

The key to simulating many replications in Excel (without an add-in) is to use a data table with any blank cell as the column input cell.

Excel Tip: How Data Tables Work

To understand this procedure, you must understand exactly how data tables work. When you create a data table, Excel takes each value in the left column of the data table (here, column A), substitutes it into the cell designated as the column input cell, recalculates the spreadsheet, and returns the output value (or values) you have requested in the top row of the data table (such as profit). It might seem silly to substitute each replication number from column A into a blank cell such as cell C23, but this part is really irrelevant. The important part is the recalculation. Each recalculation leads to a new random demand and corresponding profit, and these profits are the quantities you want to keep track of. Of course, this means that you should not freeze the quantity in cell B19 before forming the data table. The whole point of the data table is to use a different random number for each replication, and this will occur only if the random demand in row 19 is "live."

Using a Two-Way Data Table

You can carry this method one step further to see how the profit depends on the order quantity. Here you use a two-way data table with the replication number along the side and possible order quantities along the top. See Figure 15.29 and the file **Ordering Calendars - Excel**

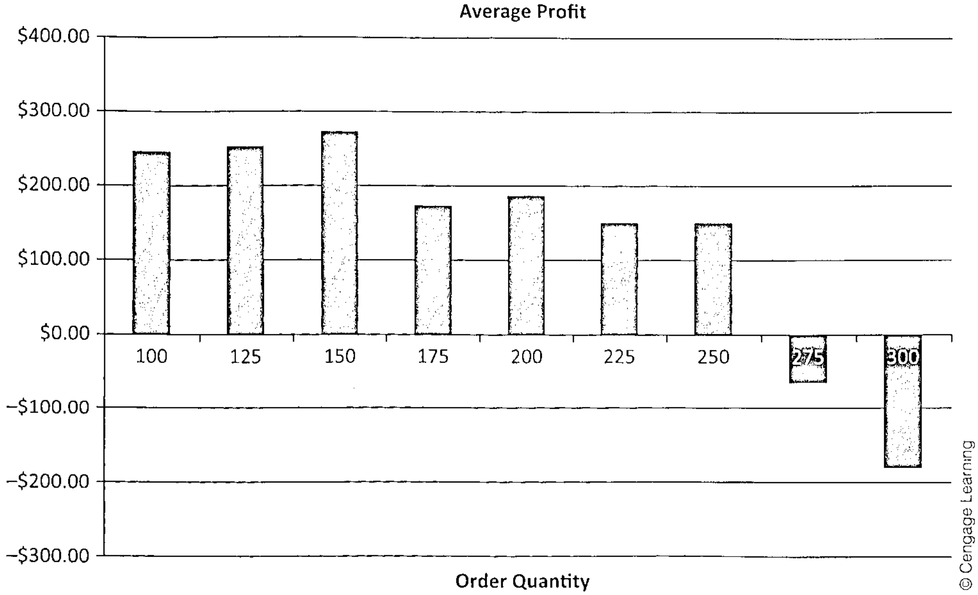
Figure 15.29 Using a Two-Way Data Table for the Simulation Model

	A	B	C	D	E	F	G	H	I	J
17	Simulation									
18		Demand	Revenue	Cost	Refund	Profit				
19		100	$1,000	$1,500	$250	–$250				
20										
21	Data table showing profit for replications with various order quantities									
22	Replication		Order quantity							
23	($250.00)	100	125	150	175	200	225	250	275	300
24	1	$250	$313	$375	–$125	$500	–375	625	125	–375
25	2	$250	$125	$375	$250	–$250	–375	–500	125	–375
26	3	$250	$313	$0	$438	$500	375	625	–250	–750
27	4	$250	$125	$375	$438	$500	0	250	–625	–750
28	5	$250	$125	$0	–$125	–$250	375	250	–625	–750
1021	998	$250	$125	$375	$438	–$250	0	–500	–625	–375
1022	999	$250	$313	$375	–$125	$125	375	–125	125	–750
1023	1000	$250	$125	$375	$438	$500	0	625	–625	–750

Only 3.xlsx. Now the data table range is A23:J1023, and the driving formula in cell A23 is again the link **=F19**. The column input cell should again be *any blank cell*, and the row input cell should be B9 (the order quantity). Each cell in the body of the data table shows a simulated profit for a particular replication and a particular order quantity, and each is based on a *different* random demand.

By averaging the numbers in each column of the data table (see row 14 in the finished version of the file), you can see which is the best order quantity. It is also helpful to construct a column chart of these averages, as in Figure 15.30. Now, however, assuming you have not frozen anything, the data table and the corresponding chart will change each time you press the F9 key. To see whether 150 is always the best order quantity, you can press the F9 key and see whether the bar above 150 continues to be the highest. (It usually is, but not always.) ■

Figure 15.30
Column Chart of Average Profits for Different Order Quantities



By now you should appreciate the usefulness of data tables in spreadsheet simulations. They allow you to take a prototype simulation and replicate its key results as often as you like. This method makes summary statistics (over the entire group of replications) and corresponding charts fairly easy to obtain. Nevertheless, it takes some work to create the data tables and charts. In the next section you will see how the @RISK add-in does a lot of this work for you.

FUNDAMENTAL INSIGHT

Decisions Based on Simulation Results

Given the emphasis in Chapter 6 on the EMV criterion for decision making under uncertainty, you might believe that the mean of an output from a simulation is the only summary measure of the output relevant for decision making. However, this is not necessarily true. When you run a simulation, you approximate the entire distribution of an output, including its mean, its standard deviation, its percentiles, and more. As a decision maker, you could base your decision on any of these summary measures, not just the mean. For example, you could focus on making the standard deviation small, making the 5th percentile large, or others. The point is that the results from a simulation provide a lot more information about an output than simply its mean, and this information can be used for decision making.

PROBLEMS

Level A

11. Suppose you own an expensive car and purchase auto insurance. This insurance has a $1000 deductible, so that if you have an accident and the damage is less than $1000, you pay for it out of your pocket. However, if the damage is greater than $1000, you pay the first $1000 and the insurance pays the rest. In the current year there is probability 0.025 that you will have an accident. If you have an accident, the damage amount is normally distributed with mean $3000 and standard deviation $750.

a. Use Excel to simulate the amount you have to pay for damages to your car. This should be a one-line simulation, so run 5000 iterations by copying it down. Then find the average amount you pay, the standard deviation of the amounts you pay, and a

95% confidence interval for the average amount you pay. (Note that many of the amounts you pay will be 0 because you have no accidents.)

b. Continue the simulation in part **a** by creating a two-way data table, where the row input is the deductible amount, varied from $500 to $2000 in multiples of $500. Now find the average amount you pay, the standard deviation of the amounts you pay, and a 95% confidence interval for the average amount you pay for each deductible amount.

c. Do you think it is reasonable to assume that damage amounts are *normally* distributed? What would you criticize about this assumption? What might you suggest instead?

12. In August of the current year, a car dealer is trying to determine how many cars of the next model year to order. Each car ordered in August costs $20,000. The demand for the dealer's next year models has the probability distribution shown in the file **P15_12.xlsx**. Each car sells for $25,000. If demand for next year's cars exceeds the number of cars ordered in August, the dealer must reorder at a cost of $22,000 per car. Excess cars can be disposed of at $17,000 per car. Use simulation to determine how many cars to order in August. For your optimal order quantity, find a 95% confidence interval for the expected profit.

13. In the Walton Bookstore example, suppose that Walton receives no money for the first 50 excess calendars returned but receives $2.50 for every calendar after the first 50 returned. Does this change the optimal order quantity?

14. A sweatshirt supplier is trying to decide how many sweatshirts to print for the upcoming NCAA basketball championships. The final four teams have emerged from the quarterfinal round, and there is now a week left until the semifinals, which are then followed in a couple of days by the finals. Each sweatshirt costs $10 to produce and sells for $25. However, in three weeks, any leftover sweatshirts will be put on sale for half price, $12.50. The supplier assumes that the demand for his sweatshirts during the next three weeks (when interest in the tournament is at its highest) has the distribution shown in the file **P15_14.xlsx**. The residual demand, after the sweatshirts have been put on sale, has the distribution also shown in this file. The supplier, being a profit maximizer, realizes that every sweatshirt sold, even at the sale price, yields a profit. However, he also realizes that any sweatshirts produced but not sold (even at the sale price) must be thrown away, resulting in a $10 loss per sweatshirt. Analyze the supplier's problem with a simulation model.

Level B

15. In the Walton Bookstore example with a discrete demand distribution, explain why an order quantity other than one of the possible demands cannot maximize the expected profit. (*Hint*: Consider an order of 190 calendars, for example. If this maximizes expected profit, then it must yield a higher expected profit than an order of 150 or 100. But then an order of 200 calendars must also yield a larger expected profit than 190 calendars. Why?)

15-5 INTRODUCTION TO @RISK

Spreadsheet simulation modeling has become extremely popular in recent years, both in the academic and corporate communities. Much of the reason for this popularity is due to simulation add-ins such as **@RISK**. There are two primary advantages to using such an add-in. First, an add-in gives you easy access to many probability distributions you might want to use in your simulation models. You already saw in Section 15-2 how the RISKDISCRETE, RISKNORMAL, and RISKTRIANG functions, among others, are easy to use and remember. Second, an add-in allows you to perform simulations much more easily than is possible with Excel alone. To replicate a simulation in Excel, you typically need to build a data table. Then you have to calculate summary statistics, such as averages, standard deviations, and percentiles, with built-in Excel functions. If you want graphs to enhance the analysis, you have to create them. In short, you have to perform a number of time-consuming steps for each simulation. Simulation add-ins such as @RISK perform much of this work automatically.

@RISK provides a number of functions for simulating from various distributions, and it takes care of all the bookkeeping in spreadsheet simulations. Excel simulations without @RISK require much more work for the user.

Although we will focus only on @RISK in this book, it is not the only simulation add-in available for Excel. Two worthy competitors are Crystal Ball, available from Oracle, and Risk Solver Platform, available from Frontline Systems, the developer of Solver. Both Crystal Ball and Risk Solver Platform have much of the same functionality as @RISK. However, the authors have a natural bias for @RISK—we have been permitted by its developer, Palisade Corporation, to provide the academic version free with this book. If it were not included, you would have to purchase it from Palisade at a fairly steep price. Indeed, Microsoft Office does not include @RISK, Crystal Ball, Risk Solver Platform, or any other simulation add-in—you must purchase them separately.

15-5a @RISK Features

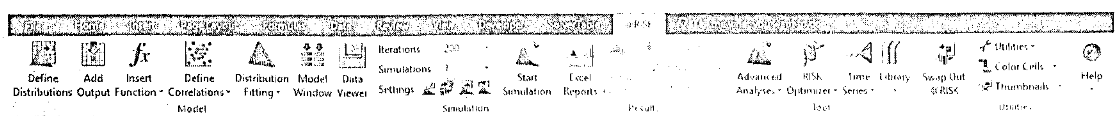
Here is an overview of some of @RISK's features. We will discuss all of these in more detail in this section.

- @RISK contains a number of functions such as RISKNORMAL and RISKDISCRETE that make it easy to generate observations from a wide variety of probability distributions. You saw some of these in Section 15-2.
- You can designate any cell or range of cells in your simulation model as *output cells*. When you run the simulation, @RISK automatically keeps summary measures (averages, standard deviations, percentiles, and others) from the values generated in these output cells across the replications. It also creates graphs such as histograms based on these values. In other words, @RISK takes care of tedious bookkeeping operations for you.
- @RISK has a special function, **RISKSIMTABLE**, that allows you to run the same simulation several times, using a different value of some key input variable each time. This input variable is often a decision variable. For example, suppose that you would like to simulate an inventory ordering policy (as in the Walton Bookstore example). Your ultimate purpose is to compare simulation outputs across a number of possible order quantities such as 100, 150, 200, 250, and 300.If you use an appropriate formula involving the RISKSIMTABLE function, the entire simulation is performed for each of these order quantities separately—with one click of a button. You can then compare the outputs to choose the best order quantity.

15-5b Loading @RISK

To build simulation models with @RISK, you need to have Excel open with @RISK added in. The first step, if you have not already done so, is to install the Palisade DecisionTools suite with its Setup program. Then you can load @RISK by clicking the Windows Start button, selecting the Programs group, selecting the Palisade DecisionTools group, and selecting @RISK. If Excel is already open, this loads @RISK inside Excel. If Excel is not yet open, this launches Excel and @RISK simultaneously.[8] After @RISK is loaded, you see an @RISK tab and the corresponding @RISK ribbon in Figure 15.31.[9]

Figure 15.31 @RISK Ribbon



© Cengage Learning

[8]We have had the best luck when we (1) close other applications we are not currently using, and (2) launch Excel and @RISK together by starting @RISK. However, you can also start @RISK *after* Excel is already running.
[9]If you have been using a previous version of @RISK, you will see some changes in the version 7.0 we are using here. However, the basic functionality and the user interface are essentially the same.

@RISK Tip: Learning the Software
When you launch @RISK, you will see a Welcome screen. (This Welcome screen is also available at any time from the @RISK Help menu.) Among other things, this Welcome screen has a Quick Start link, which guides beginners through the basic features of @RISK. It also has a Guided Tours link to a series of videos on @RISK's basic to advanced features. All of these videos were created for Palisade by Albright. (He also developed a number of similar videos that accompany this book.)

15-5c @RISK Models with a Single Random Input Variable

The majority of the work (and thinking) goes into developing the model. Setting up @RISK and then running it are relatively easy.

In the remainder of this section we will illustrate some of @RISK's functionality by revisiting the Walton Bookstore example. The next chapter demonstrates the use of @RISK in a number of interesting simulation models. Throughout this discussion, you should keep one very important idea in mind. The development of a simulation model is basically a two-step procedure. The first step is to build the model itself. This step requires you to enter all of the logic that transforms inputs (including @RISK functions such as RISKDISCRETE) into outputs (such as profit). This is where most of the work and thinking go, exactly as in models from previous chapters, and @RISK cannot do this for you. It is *your* job to enter the formulas that link inputs to outputs appropriately. However, once this logic has been incorporated, @RISK takes over in the second step. It automatically replicates your model, with different random numbers on each replication, and it reports any summary measures that you request in tabular or graphical form. Therefore, @RISK greatly decreases the amount of busy work you need to do, but it is not a magic bullet.

We begin by analyzing an example with a single random input variable.

EXAMPLE 15.3 Using @RISK at Walton Bookstore

Recall that Walton Bookstore buys calendars for $7.50, sells them at the regular price of $10, and gets a refund of $2.50 for all calendars that cannot be sold. In contrast to Example 15.2, we assume now that Walton estimates a triangular probability distribution for demand, where the minimum, most likely, and maximum values of demand are 100, 175, and 300, respectively. The company wants to use this probability distribution, together with @RISK, to simulate the profit for any particular order quantity, with the ultimate goal of finding the best order quantity.

This is the same Walton Bookstore model as before, except that a triangular distribution for demand is used.

Objective To learn about @RISK's basic functionality by revisiting the Walton Bookstore problem.

Where Do the Numbers Come From?

The monetary values are the same as before. The parameters of the triangular distribution of demand are probably Walton's best subjective estimates, possibly guided by its experience with previous calendars. As in many simulation examples, the triangular distribution is chosen for simplicity. In this case, the manager would need to estimate only three quantities: the minimum possible demand, the maximum possible demand, and the most likely demand.

Solution

We use this example to illustrate important features of @RISK. We first show how it helps you to implement an appropriate input probability distribution for demand. Then we show

how it can be used to build a simulation model for a specific order quantity and generate outputs from this model. Finally, we show how the RISKSIMTABLE function can be used to simultaneously generate outputs from several order quantities so that you can choose the optimal order quantity.

Developing the Simulation Model

Developing the Walton Model with @RISK

The spreadsheet model for profit is essentially the same model developed previously *without* @RISK, as shown in Figure 15.32. (See the file **Ordering Calendars - @RISK.xlsx.**) There are only a few new things to be aware of.

@RISK Tip: Settings When Opening a Workbook

When you open a workbook with an @RISK model, such as those that accompany this chapter, you might be asked whether you want to change the current @RISK settings to match those stored in the workbook. You should generally click Yes. This changes settings, such as the number of iterations, to those you stored previously with the workbook instead of using @RISK default settings.

Figure 15.32 Simulation Model with a Fixed Order Quantity

	A	B	C	D	E	F	G	H	I	J
1	**Simulation of Walton's Bookstore using @RISK**							**Range names used:**		
2								Order_quantity	=Model!$B$9	
3	**Cost data**			**Demand distribution - triangular**				Profit	=Model!$F$13	
4	Unit cost	$7.50		Minimum	100			Unit_cost	=Model!$B$4	
5	Unit price	$10.00		Most likely	175			Unit_price	=Model!$B$5	
6	Unit refund	$2.50		Maximum	300			Unit_refund	=Model!$B$6	
7										
8	**Decision variable**									
9	Order quantity	200								
10										
11	**Simulation**									
12		Demand	Revenue	Cost	Refund	Profit				
13		192	$1,920	$1,500	$20	$440				
14										
15	**Summary measures of profit from @RISK - based on 1000 iterations**									
16	Minimum	–$242.50								
17	Maximum	$500.00								
18	Average	$337.48								
19	Standard deviation	$189.09								
20	5th percentile	–$47.50								
21	95th percentile	$500.00								
22	P(profit<=300)	0.360								
23	P(profit>400)	0.514								

1. **Input distribution.** To generate a random demand, enter the formula

   =ROUND(RISKTRIANG(E4,E5,E6),0)

   in cell B13 for the random demand. This uses the RISKTRIANG function to generate a demand from the triangular distribution. (As before, our convention is to color random input cells green.) Excel's ROUND function is used to round demand to the nearest integer. Recall from the discussion in Section 15-3 that Excel has no built-in functions to generate random numbers from a triangular distribution, but this is easy with @RISK.
2. **Output cell.** When the simulation runs, you want @RISK to keep track of profit. In @RISK's terminology, you need to designate the Profit cell, F13, as an *output cell*. To do this, select cell F13 and then click the Add Output button on the @RISK ribbon.

(See Figure 15.31.) This adds **RISKOUTPUT(*"label"*)+** to the cell's formula. (Here, "label" is a label that @RISK uses for its reports. In this case it makes sense to use "Profit" as the label.) The formula in cell F13 changes from

=C13+E13-D13

to

=RISKOUTPUT("Profit")+C13+E13-D13

The RISKOUTPUT function indicates that a cell is an output cell, so that @RISK will keep track of its values throughout the simulation.

The plus sign following **RISKOUTPUT** is simply @RISK's way of indicating that you want to keep track of the value in this cell (for reporting reasons) as the simulation progresses. Any number of cells can be designated in this way as output cells. They are typically the "bottom line" values of primary interest. Our convention is to color such cells gray for emphasis.

3. **Summary functions (optional).** There are several places where you can store @ RISK results. One of these is to use @RISK statistical functions to place results in your model worksheet. @RISK provides several functions for summarizing output values. Some of these are illustrated in the range B16:B23 of Figure 15.32. They contain the formulas

These @RISK summary functions allow you to show simulation results on the same sheet as the model. However, they are totally optional.

=RISKMIN(F13)

=RISKMAX(F13)

=RISKMEAN(F13)

=RISKSTDDEV(F13)

=RISKPERCENTILE(F13,0.05)

=RISKPERCENTILE(F13,0.95)

=RISKTARGET(F13,300)

and

=1-RISKTARGET(F13,400)

The values in these cells are not meaningful until you run the simulation (so do not be alarmed if they contain errors when you open the file). However, once the simulation runs, these formulas capture summary statistics of profit. For example, RISKMEAN calculates the average of the 1000 simulated profits, RISKPERCENTILE finds the value such that the specified percentage of simulated profits are less than or equal to this value, and RISKTARGET finds the percentage of simulated profits less than or equal to the specified value. Although these same summary statistics also appear in other @RISK reports, it is handy to have them in the same worksheet as the model. (You can find a list of all @RISK statistical functions from the Simulation Result group in the Insert Function dropdown list on the @RISK ribbon.)

New Color Coding Feature

@RISK Feature: Color Coding

One very handy @RISK feature is its optional color coding. This option is in the Utilities group of the @RISK ribbon. It is a toggle. If it is toggled off, you see our blue/red/gray/ green coloring. If it is toggled on, you see @RISK's color coding: blue for random input cells, red for output cells, green for statistical functions, and yellow for decision cells (for RISKOptimizer models). @RISK even allows you to change the coloring scheme if you prefer.

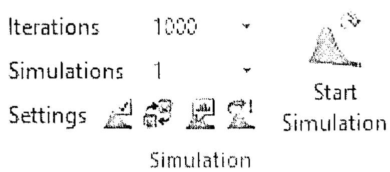
Running the Simulation

After you develop the model, the rest is straightforward. The procedure is always the same: (1) specify simulation settings, (2) run the simulation, and (3) examine the results.

Simulation and Application Settings in @RISK

1. **Simulation settings.** You must first choose some simulation settings. To do so, the buttons on the left in the Simulation group (see Figure 15.33) are useful. We typically do the following:
   - Set Iterations to a number such as 1000. (@RISK calls replications "iterations.") Any number can be used, but because the academic version of @RISK allows only 1000 uninterrupted iterations, we typically choose 1000.
   - Set Simulations to 1. In a later section, we will explain why you might want to request multiple simulations.
   - The "dice" button is a toggle for what appears in your worksheet. If it is colored (Random), all random cells appear random (they change when you press the F9 key). If it is white (Static), only the *means* appear in random input cells and the F9 key has no effect. We tend to prefer the Random setting, but this setting is irrelevant when you run the simulation.
   - Many more settings are available by clicking the Simulation Settings button to the left of the "dice" button, but the ones we mentioned should suffice. In addition, more permanent settings can be chosen from Application Settings in the Utilities dropdown list on the @RISK ribbon. You can experiment with these, but the only one we like to change is the Place Reports In setting in the Reports group. The default is to place reports in a new workbook. If you like the reports to be in the same workbook as your model, you can change this setting to Active Workbook.

Figure 15.33
Simulation Group on @RISK Ribbon



© Cengage Learning

Leave Latin Hyper-cube sampling on. It produces more accurate results.

Latin Hypercube vs Monte Carlo Sampling

@RISK Technical Issues: Latin Hypercube Sampling and Mersenne Twister Generator

Two settings you shouldn't change are the Sampling Type and Generator settings (available from the Simulation Settings button and then the Sampling tab). They should remain at the default Latin Hypercube and Mersenne Twister settings. The Mersenne Twister is one of many algorithms for generating random numbers, and it has been shown to have very good statistical properties. (Not all random number generators do.) ***Latin Hypercube sampling*** *is a more efficient way of sampling than the other option (Monte Carlo) because it produces a more accurate estimate of the output distribution. In fact, we were surprised how accurate it is. In repeated runs of this model, always using different random numbers, we virtually always got a mean profit within a few pennies of $337.50. It turns out that this is the true mean profit for this input distribution of demand. Amazingly, simulation estimates it correctly—almost exactly—on virtually every run. However, this means that a confidence interval for the mean, based on @RISK's outputs and the usual confidence interval formula (which assumes Monte Carlo sampling), is much wider (more pessimistic) than it should be. Therefore, we do not even calculate such confidence intervals from here on. However, it is not impossible. The accompanying video explains a* *method called Batch Means for calculating confidence intervals when Latin Hypercube sampling is used.*

2. **Run the simulation.** To run the simulation, click the Start Simulation button on the @RISK ribbon. When you do so, @RISK repeatedly generates a random number for each random input cell, recalculates the worksheet, and keeps track of all output cell values. You can watch the progress at the bottom left of the screen. Also, if the Automatically Show Output Graph button (to the right of the dice button) is toggled to colored, you will see a histogram of the currently selected input or output cell being built as the simulation runs. If you find this annoying, you can toggle this button to white.
3. **Examine the results.** The big questions are (1) which results you want and (2) where you want them. @RISK provides a lot of possibilities, and we mention the most frequently used.
   - You can ask for summary measures in your model worksheet by using the @RISK statistical functions, such as RISKMEAN, discussed earlier.
   - The quickest way to get results is to select an input or output cell (we chose the profit cell, F13) and then click the Browse Results button in the Results group of the @RISK ribbon. (See Figure 15.34.) This provides an interactive graph of the selected value, as shown in Figure 15.35. You can move the "sliders" (the two vertical bars) on this graph to see probabilities of various outcomes. Note that the window you see from Browse Results is temporary—it goes away when you click Close. You can make a permanent copy of the chart by clicking the third button from the left (see the bottom of Figure 15.35) and choosing one of the copy options.

For a quick graph of the distribution of an output or input, select the output or input cell and click @RISK's Browse Results button.

Figure 15.34
Results Group on @RISK Ribbon

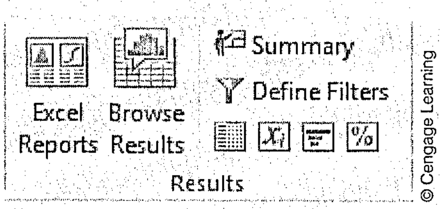
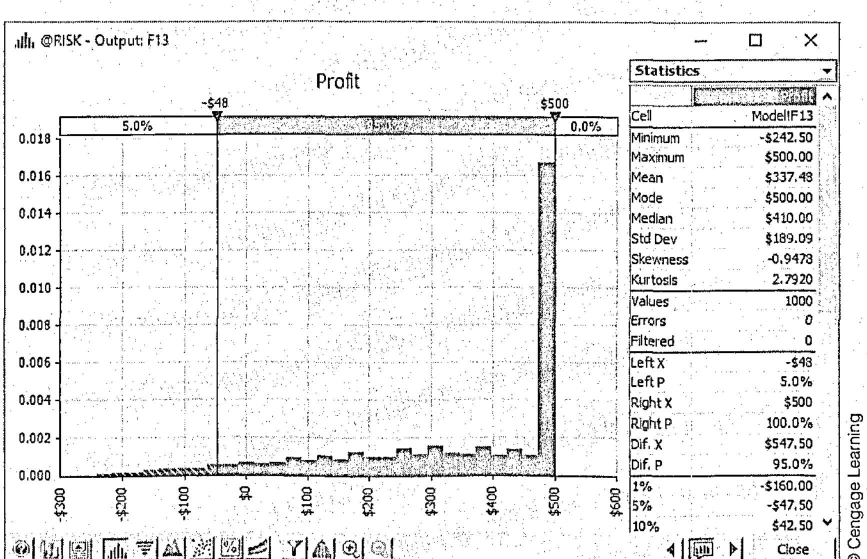


Figure 15.35
Interactive Graph of Profit Distribution



@RISK Tip: Graph Type

Once you get an @RISK graph such as the one in Figure 15.35, you can display it in several ways by clicking the fourth button at the bottom of the graph window. In particular, if your graph doesn't look quite like the ones shown here, try changing from Discrete Probability to Probability Density or vice versa.

@RISK Tip: Percentiles Displayed on Graphs

When we displayed the graph in Figure 15.35 the first time, it had the right slider on 500 but showed 5% to the right of it. By default, @RISK puts the sliders at the 5th and 95th percentiles, so that 5% is on either side of them. For this example, 500 is indeed the 95th percentile (why?), but the picture is a bit misleading because there is no chance of a profit greater than 500. When we manually moved the right slider away from 500 and back again, it displayed as in Figure 15.35, correctly indicating that there is no probability to the right of 500.

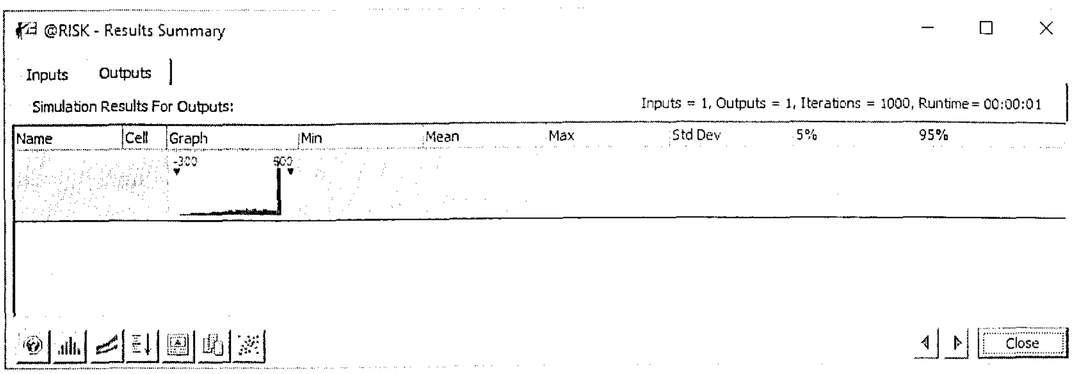
@RISK Tip: Saving Graphs and Tables

When you run a simulation with @RISK and then save your file, it asks whether you want to save your graphs and tables. We suggest that you save them. This makes your file slightly larger, but when you reopen it, the temporary graphs and tables, such as the histogram in Figure 15.35, are still available. Otherwise, you will have to rerun the simulation.

For a quick (and customizable) report of the results, click @RISK's Summary button.

- You can click the Summary button (again, see Figure 15.34) to see the window in Figure 15.36 with the summary measures for Profit. In general, this report shows the summary for *all* designated inputs and outputs. By default, this Results Summary window shows a mini graph for each output and a number of numerical summary measures. It is also easy to customize. If you right-click anywhere on this table and choose Columns for Table, you can check or uncheck various options. For most of the later screenshots in this book, we elected *not* to show the Error column, but instead to show the standard deviation column.

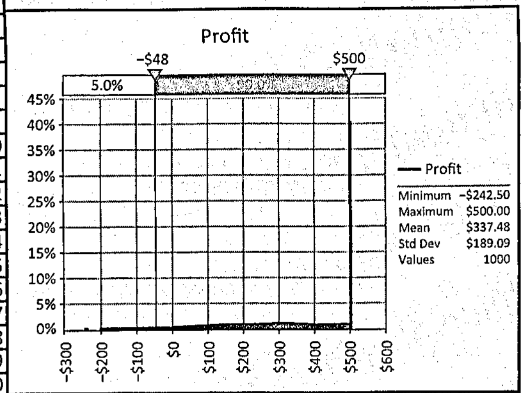
Figure 15.36 Summary Table of Profit Output

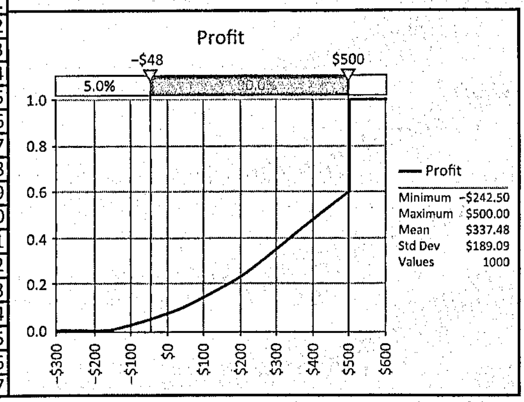


If you want permanent copies of the simulation results, click on @RISK's Excel Reports buttons and check the reports you want. They will be placed in new worksheets.

- You can click the Excel Reports button (again, see Figure 15.34) to choose from a number of reports that are placed on new worksheets. This is a good option if you want permanent (but non-interactive) copies of reports in your workbook. As an example, Figure 15.37 shows part of the Quick Reports option you can request. It has the same information as the summary report in Figure 15.36, plus more.

Figure 15.37
@RISK Quick Report





Simulation Summary Information	
Workbook Name	Ordering Calendars - @RI
Number of Simulations	1
Number of Iterations	1000
Number of Inputs	1
Number of Outputs	1
Sampling Type	Latin Hypercube
Simulation Start Time	8/25/2015 11:58
Simulation Duration	00:00:01
Random # Generator	Mersenne Twister
Random Seed	1

Summary Statistics for Profit			
Statistics		Percentile	
Minimum	-$243	5%	-$48
Maximum	$500	10%	$43
Mean	$337	15%	$103
Std Dev	$189	20%	$163
Variance	35754.13508	25%	$208
Skewness	-0.947766786	30%	$253
Kurtosis	2.792008113	35%	$290
Median	$410	40%	$328
Mode	$500	45%	$373
Left X	-$48	50%	$410
Left P	5%	55%	$455
Right X	$500	60%	$500
Right P	95%	65%	$500
Diff X	$548	70%	$500
Diff P	90%	75%	$500
#Errors	0	80%	$500
Filter Min	Off	85%	$500
Filter Max	Off	90%	$500
#Filtered	0	95%	$500

Discussion of the Simulation Results

The strength of @RISK is that it keeps track of any outputs you designate and then allows you to show the corresponding results as graphs or tables, in temporary windows or in permanent worksheets. As you have seen, @RISK provides several options for displaying results, and we encourage you to explore the possibilities. However, don't lose sight of the overall goal: to see how outputs vary as random inputs vary, and to generate reports that tell the story most effectively. For this particular example, the results in Figures 15.32, 15.35, 15.36, and 15.37 allow you to conclude the following:

- The smallest simulated profit (out of 1000) was –$242.50, the largest was $500, the average was $337.48, and the standard deviation of the 1000 profits was $189.09. Of all simulated profits, 5% were –$47.50 or below, 95% were $500 or above, 36% were less than or equal to $300, and 51.4% were larger than $400. (See Figure 15.32. These results are also available from the summary table in Figure 15.36 or the quick report in Figure 15.37.)
- The profit distribution for this particular order quantity is extremely skewed to the left, with a large bar at $500. (See Figure 15.35.) Do you see why? It is because profit is exactly $500 if demand is greater than or equal to the order quantity, 200. In other words, the probability that profit is $500 equals the probability that demand is at least 200. (This probability is 0.4.) Lower demands result in decreasing profits, which explains the gradual decline in the histogram from right to left.

Using RISKSIMTABLE

Walton's ultimate goal is to choose an order quantity that provides a large average profit. You could rerun the simulation model several times, each time with a different order quantity in the order quantity cell, and compare the results. However, this has two drawbacks. First, it takes a lot of time and work. The second drawback is more subtle. Each time you run the simulation, you get a *different* set of random demands. Therefore, one of the order quantities could win the contest just by luck. For a fairer comparison, it is better to test each order quantity on the *same* set of random demands.

The RISKSIMTABLE function allows you to run several simulations at once—one for each value of some variable (often a decision variable).

The RISKSIMTABLE function in @RISK enables you to obtain a fair comparison quickly and easily. This function is illustrated in Figure 15.38. (See the file **Ordering Calendars - RiskSimtable.xlsx**.) There are two modifications to the previous model. First, the order quantities to test are listed in row 9. (We chose these as representative order quantities. You could change, or add to, this list.) Second, instead of entering a *number* in cell B9, you enter the *formula*

=RISKSIMTABLE(D9:H9)

Note that the list does not need to be entered in the spreadsheet (although it is a good idea to do so). You could instead enter the formula

=RISKSIMTABLE({150,175,200,225,250})

where the list of numbers must be enclosed in curly brackets. In either case, the worksheet displays the first member of the list, 150, and the corresponding calculations for this first order quantity. However, the model is now set up to run the simulation for *all* order quantities in the list.

Figure 15.38 Model with a RISKSIMTABLE Function

	A	B	C	D	E	F	G	H	I	J	K
1	Simulation of Walton's Bookstore using @RISK								Range names used:		
2									Order_quantity	=Model!$B$9	
3	Cost data			Demand distribution - triangular					Profit	=Model!$F$13	
4	Unit cost	$7.50		Minimum	100				Unit_cost	=Model!$B$4	
5	Unit price	$10.00		Most likely	175				Unit_price	=Model!$B$5	
6	Unit refund	$2.50		Maximum	300				Unit_refund	=Model!$B$6	
7											
8	Decision variable			Order quantities to try							
9	Order quantity	150		150	175	200	225	250			
10											
11	Simulated quantities										
12		Demand	Revenue	Cost	Refund	Profit					
13		178	$1,500	$1,125	$0	$375					
14											
15	Summary measures of profit from @RISK - based on 1000 iterations for each simulation										
16	Simulation	1	2	3	4	5					
17	Order quantity	150	175	200	225	250					
18	Minimum	$22.50	–$102.50	–$227.50	–$352.50	–$477.50					
19	Maximum	$375.00	$437.50	$500.00	$562.50	$625.00					
20	Average	$354.20	$367.22	$337.51	$270.32	$175.02					
21	Standard deviation	$58.94	$121.81	$189.01	$247.01	$286.95					
22	5th percentile	$202.50	$77.50	–$47.50	–$172.50	–$297.50					
23	95th percentile	$375.00	$437.50	$500.00	$562.50	$625.00					

To implement this, only one setting needs to be changed. As before, enter 1000 for the number of iterations, but also enter 5 for the number of simulations. @RISK then runs five simulations of 1000 iterations each, one simulation for each order quantity in the list, and it uses the *same* 1000 random demands for each simulation. This provides a fair comparison.

@RISK Function: RISKSIMTABLE

To run several simulations all at once, enter the formula ***=RISKSIMTABLE(InputRange)*** *in any cell. Here, InputRange refers to a list of the values to be simulated, such as various order quantities. Before running the simulation, make sure the number of simulations is set to the number of values in the InputRange list.*

You can again get results from the simulation in various ways. Here are some possibilities.

- You can enter the same @RISK statistical functions in cells in the model work sheet, as shown in rows 18–23 of Figure 15.38. The trick is to realize that each such function has an optional last argument that specifies the simulation number. For example, the formulas in cells C20 and C22 are

  =RISKMEAN($F$13,C16)

  and

  =RISKPERCENTILE($F$13,0.05,C16)

  Remember that the results in these cells are meaningless (or show up as errors) until you run the simulation.
- You can select the profit cell and click the Browse Results button to see a histogram of profits, as shown in Figure 15.39. By default, the histogram shown is for the *first* simulation, where the order quantity is 150. However, if you click the red histogram button with the pound sign, you can select any of the simulations. As an example, Figure 15.40 shows the histogram of profits for simulation 5, where the order quantity is 250. (Do you see why these two histograms are so different? When the order quantity is 150, there is a high probability of selling out, so the spike on the right is large. But the probability of selling out with an order quantity of 250 is much lower, so its spike at the right is much less dominant.)
- You can click the Summary button to get the results from all simulations shown in Figure 15.41. (These results match those in Figure 15.38.)
- You can click the Excel Reports button to get any of a number of reports on permanent worksheets. Again, Quick Reports is a good choice. This produces several graphs and summary measures for each simulation, each on a different worksheet. This provides a lot of information with almost no work.

Figure 15.39
Histogram of Profit with Order Quantity 150

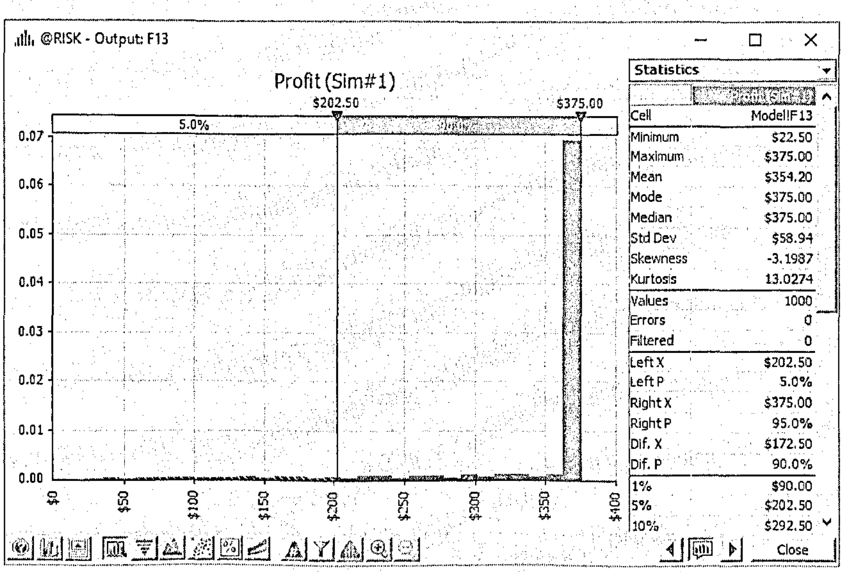


Figure 15.40 Histogram of Profit with Order Quantity 250

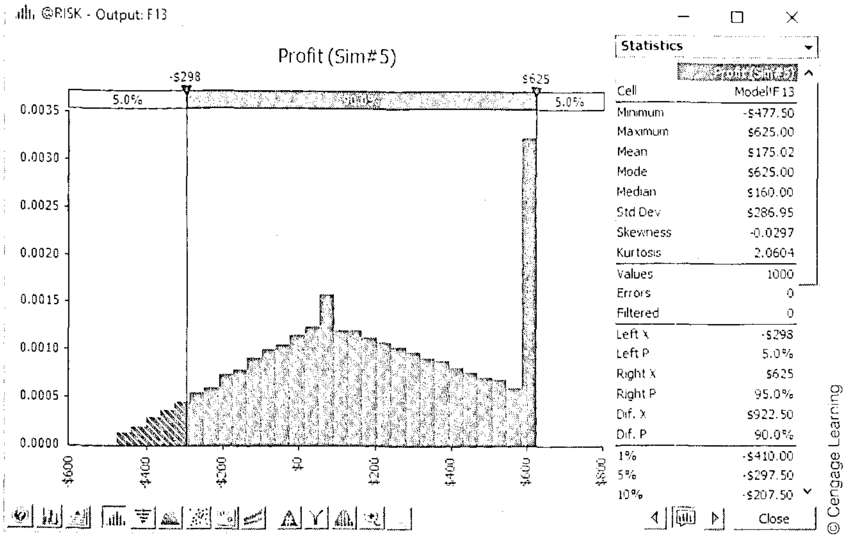
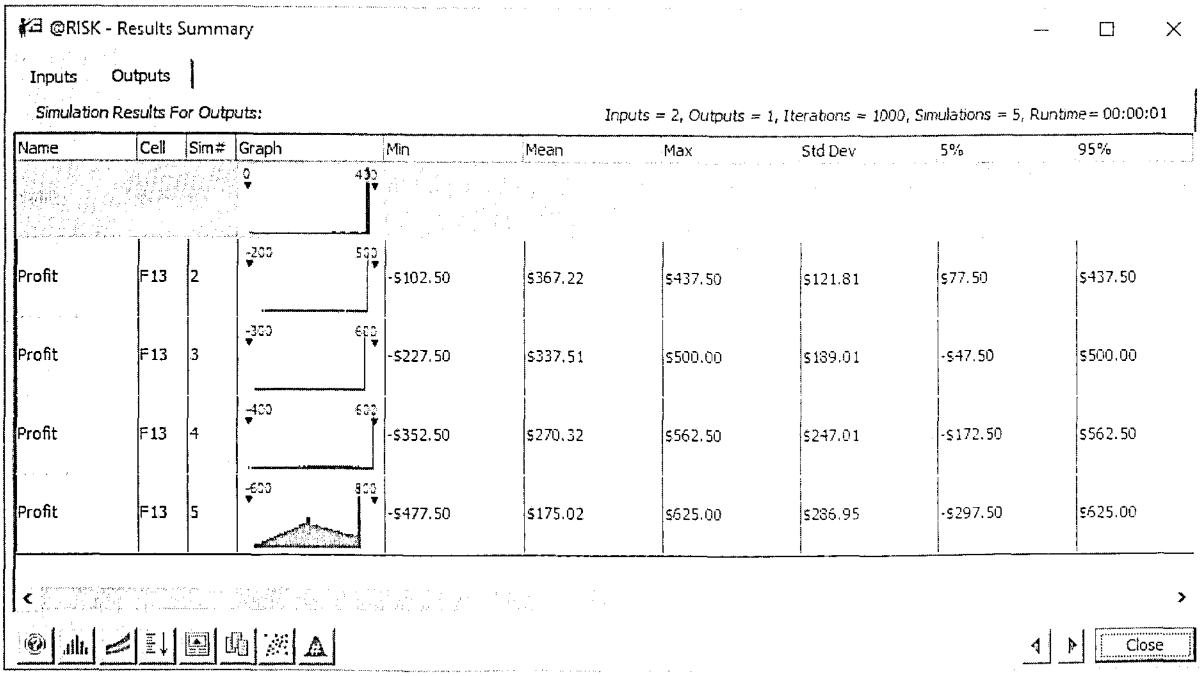


Figure 15.41 Summary Report for All Five Simulations



Name	Cell	Sim#	Graph	Min	Mean	Max	Std Dev	5%	95%
Profit	F13	2	-200 500	-$102.50	$367.22	$437.50	$121.81	$77.50	$437.50
Profit	F13	3	-300 600	-$227.50	$337.51	$500.00	$189.01	-$47.50	$500.00
Profit	F13	4	-400 600	-$352.50	$270.32	$562.50	$247.01	-$172.50	$562.50
Profit	F13	5	-600 800	-$477.50	$175.02	$625.00	$286.95	-$297.50	$625.00

For this particular example, the results in Figures 15.38–15.41 are illuminating. You can see that an order quantity of 175 provides the largest *mean* profit. However, is this necessarily the optimal order quantity? This depends on the company's attitude toward risk. Certainly, larger order quantities incur more risk (their histograms are more spread out, their 5th and 95th percentiles are more extreme), but they also have more upside potential. On the other hand, a smaller order quantity, while having a somewhat smaller mean, might be preferable because of less variability. It is *not* an easy choice, but at least the simulation results provide plenty of information for making the decision. ◘

15-5d Some Limitations of @RISK

The academic version of @RISK included with the book has some limitations you should be aware of. (The commercial version of @RISK doesn't have these limitations. Also, the exact limitations could change as newer academic versions become available.)

- The simulation model must be contained in a single workbook with at most four worksheets, and each worksheet is limited to 300 rows and 100 columns.
- The number of @RISK input probability distribution functions, such as RISKNORMAL, is limited to 100.
- The number of unattended iterations is limited to 1000. You can request more than 1000, but you have to click a button after each 1000 iterations.
- All @RISK graphs contain a watermark.
- The Distribution Fitting tool can handle only 250 observations.

To avoid potential problems, close all other workbooks when running an @RISK model.

The first limitation shouldn't cause problems, at least not for the fairly small models discussed in this book. However, we strongly urge you to close all other workbooks when you are running an @RISK simulation model, *especially* if they also contain @RISK functions. @RISK does a lot of recalculation, both in your active worksheet and in all other worksheets or workbooks that are open. So if you are experiencing extremely slow simulations, this is probably the reason.

The second limitation can be a problem, especially in multiperiod problems. For example, if you are simulating 52 weeks of a year, and each week requires two random inputs, you are already over the 100-function limit. One way to get around this is to use built-in Excel functions for random inputs rather than @RISK functions whenever possible. For example, if you want to simulate the flip of a fair coin, the formula **=IF(RAND()<0.5,"Heads","Tails")** works just as well as the formula **=IF(RISKUNIFORM(0,1)<0.5,"Heads","Tails")**, but the former doesn't count against the 100-function limit.

15-5e @RISK Models with Several Random Input Variables

We conclude this section with another modification of the Walton Bookstore example. To this point, there has been a single random variable, demand. Often there are several random variables, each reflecting some uncertainty, and you want to include each of these in the simulation model. Example 15.4 illustrates how this can be done, and it also illustrates a very useful feature of @RISK, its sensitivity analysis.

EXAMPLE 15.4 ADDITIONAL UNCERTAINTY AT WALTON BOOKSTORE

As in the previous Walton Bookstore example, Walton needs to place an order for next year's calendar. We continue to assume that the calendars sell for \$10 and customer demand for the calendars at this price is triangularly distributed with minimum value, most likely value, and maximum value equal to 100, 175, and 300. However, there are now two other sources of uncertainty. First, the maximum number of calendars Walton's supplier can supply is uncertain and is modeled with a triangular distribution. Its parameters are 125 (minimum), 200 (most likely), and 250 (maximum). Once Walton places an order, the supplier will charge \$7.50 per calendar *if* he can supply the entire Walton order. Otherwise, he will charge only \$7.25 per calendar. Second, unsold calendars can no longer be returned to the supplier for a refund. Instead, Walton will put them on sale for \$5 apiece after January 1. At that price, Walton believes the demand for leftover calendars is triangularly distributed with parameters 0, 50, and 75. Any calendars *still* left over, say, after March 1, will be thrown away. Walton again wants to use simulation to analyze the resulting profit for various order quantities.

Objective To develop and analyze a simulation model with multiple sources of uncertainty using @RISK, and to introduce @RISK's sensitivity analysis features.

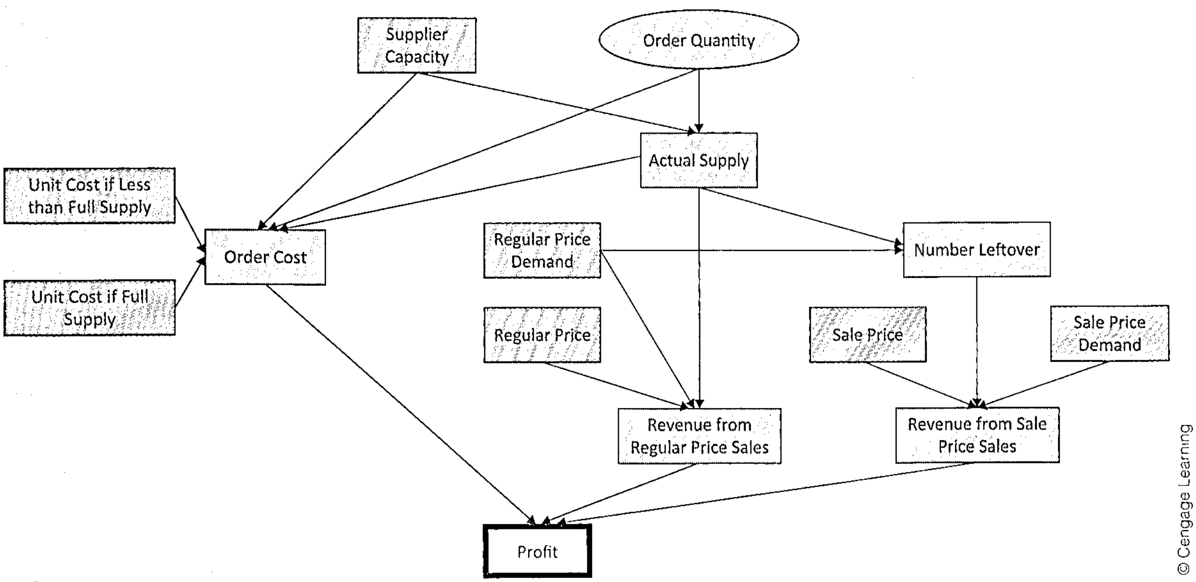
Where Do the Numbers Come From?

As in Example 15.3, the monetary values are straightforward, and the parameters of the triangular distributions are probably educated guesses, possibly based on experience with previous calendars.

Solution

The variables for this model, including the three sources of uncertainty, are shown in Figure 15.42. (See the file **Ordering Calendars - More Uncertainty Big Picture.xlsx.**) Using this as a guide, the first step, as always, is to develop the model. Then you can run the simulation with @RISK and examine the results.

Figure 15.42 Big Picture for Ordering Model with More Uncertainty



The Walton Model with Multiple Uncertain Inputs

Developing the Simulation Model

The completed model is shown in Figure 15.43. (See the file **Ordering Calendars - More Uncertainty.xlsx.**) The model itself requires a bit more logic than the previous Walton model. It can be developed with the following steps.

1. **Random inputs.** There are three random inputs in this model: the maximum supply the supplier can provide Walton, the customer demand when the selling price is $10, and the customer demand for sale-price calendars. Generate these in cells B14, E14, and H14 (using the ROUND function to obtain integers) with the formulas

   =ROUND(RISKTRIANG(I5,I6,I7),0)

   =ROUND(RISKTRIANG (E5,E6,E7),0)

Figure 15.43 @RISK Simulation Model with Three Random Inputs

	A	B	C	D	E	F	G	H	I	J
1	Simulation of Walton's Bookstore with more uncertainty									
2										
3	Cost data			Demand distribution: triangular						
4	Unit cost 1	$7.50			Regular price	Sale price		Supply distribution: triangular		
5	Unit cost 2	$7.25		Minimum	100	0		Minimum	125	
6	Regular price	$10.00		Most likely	175	50		Most likely	200	
7	Sale price	$5.00		Maximum	300	75		Maximum	250	
8										
9	Decision variable			Order quantities to try						
10	Order quantity	150		150	175	200	225	250		
11										
12	Simulated quantities				At regular price			At sale price		
13		Maximum supply	Actual supply	Cost	Demand	Revenue	Left over	Demand	Revenue	Profit
14		177	150	$1,125	139	$1,390	11	57	$55	$320
15										
16	Summary measures of profit from @RISK - based on 1000 iterations for each simulation							Range names used:		
17	Simulation	1	2	3	4	5		Order_quantity	=Model!$B$10	
18	Order quantity	150	175	200	225	250		Regular_price	=Model!$B$6	
19	Minimum	$5.00	−$82.50	−$195.00	−$377.50	−$372.00		Sale_price	=Model!$B$7	
20	Maximum	$409.75	$478.50	$547.25	$616.00	$665.50		Unit_cost_1	=Model!$B$4	
21	Average	$361.43	$389.48	$395.41	$395.58	$398.06		Unit_cost_2	=Model!$B$5	
22	Standard deviation	$43.11	$94.37	$148.47	$178.22	$180.45				
23	5th percentile	$260.00	$167.50	$46.50	-$22.75	-$33.25				
24	95th percentile	$375.00	$456.50	$528.00	$572.00	$585.75				

and

=ROUND(RISKTRIANG (F5,F6,F7),0)

Note that the formula in cell H14 generates the random *potential* demand for calendars at the sale price, even though there might not be any calendars left to put on sale.

2. **Actual supply.** The number of calendars supplied to Walton is the smaller of the number ordered and the maximum the supplier is able to supply. Calculate this value in cell C14 with the formula

   =MIN(B14,Order_quantity)

3. **Order cost.** Walton gets the reduced price, $7.25, if the supplier cannot supply the entire order. Otherwise, Walton must pay $7.50 per calendar. Therefore, calculate the total order cost in cell D14 with the formula (using the obvious range names)

   =IF(B14>=Order_quantity,Unit_cost_1,Unit_cost_2)*C14

4. **Other quantities.** The rest of the model is straightforward. Calculate the revenue from regular-price sales in cell F14 with the formula

   =Regular_price*MIN(C14,E14)

   Calculate the number left over after regular-price sales in cell G14 with the formula

   =MAX(C14-E14,0)

   Calculate the revenue from sale-price sales in cell I14 with the formula

   =Sale_price*MIN(G14,H14)

   Finally, calculate profit in cell J14 with the formula

=F14+I14-D14

Then designate this cell as an @RISK output cell. If you like, you can also designate other cells (the revenue cells, for example) as output cells.

5. **Order quantities.** As before, enter the following RISKSIMTABLE formula in cell B10 so that Walton can try different order quantities:

=RISKSIMTABLE(D10:H10)

Running the Simulation

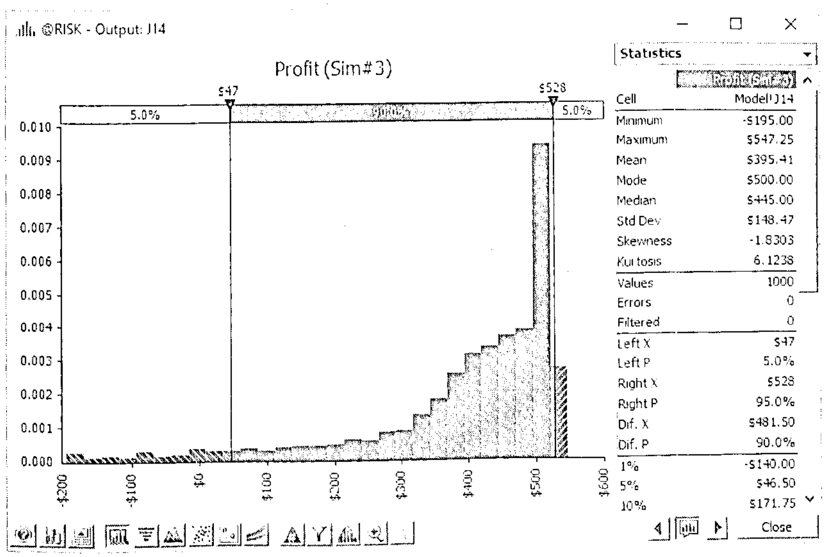
On each iteration, @RISK generates a new set of random inputs and calculates the corresponding output(s).

As usual, the next steps are to specify the simulation settings (we chose 1000 iterations and 5 simulations), and run the simulation. It is important to realize what @RISK does when it runs a simulation when there are several random input cells. In each iteration, @RISK generates a random value for each input variable *independently*. In this example, it generates a maximum supply in cell B14 from one triangular distribution, it generates a regular-price demand in cell E14 from another triangular distribution, and it generates a sale-price demand in cell H14 from a third triangular distribution. With these input values, it then calculates profit. For each order quantity, it then iterates this procedure 1000 times and keeps track of the corresponding profits.[10]

Discussion of the Simulation Results

Selected results are listed in Figure 15.43 (at the bottom), and the profit histogram for an order quantity of 200 is shown in Figure 15.44. (The histograms for the other order quantities are similar to what you have seen before, with more skewness to the left and a larger spike to the right as the order quantity decreases.) For this particular order quantity, the results indicate an average profit of $395.41, a 5th percentile of $46.50, a 95th percentile of $528, and a distribution of profits that is again skewed to the left.

Figure 15.44
Histogram of Simulated Profits for Order Quantity 200

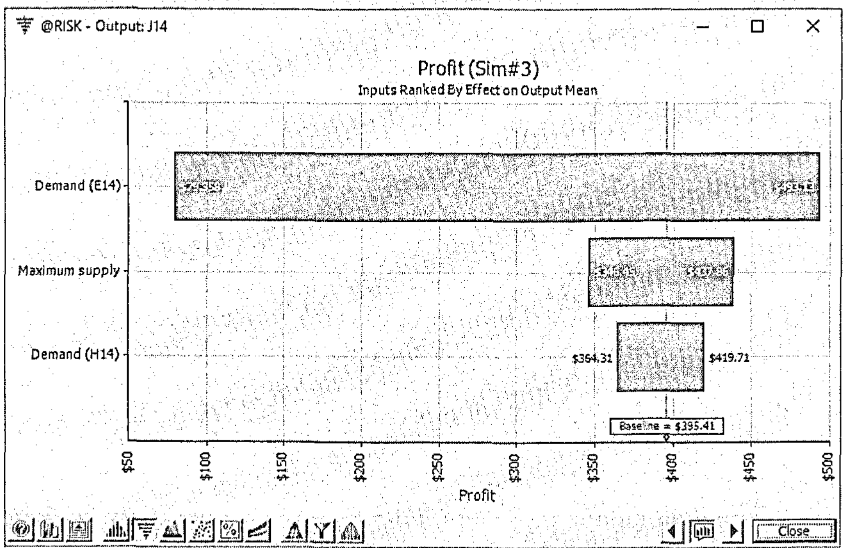


[10]It is also possible to *correlate* the inputs, as we demonstrate in the next section.

Sensitivity Analysis

We now demonstrate a feature of @RISK that is particularly useful when there are several random input cells. This feature lets you see which of these inputs has the most effect on an output cell. To perform this analysis, select the profit cell, J14, and click the Browse Results button. You will see a histogram of profit, as we have already discussed, with a number of buttons at the bottom of the window. Click the red button with the pound sign to select a simulation. We chose #3, where the order quantity is 200. Then click the "tornado" button (the fifth button from the left) and choose Change in Output Mean. This produces the graph in Figure 15.45. (The Regression and Correlation options produce similar results, and you can choose any of them if you like.)

Figure 15.45
Tornado Graph for Sensitivity Analysis



A tornado graph indicates which of the random inputs have large effects on an output.

This figure shows graphically and numerically how each of the random inputs affects profit: the longer the bar, the stronger the relationship between that input and profit. Specifically, each bar shows how the mean profit varies as each input varies over its range (and the other inputs are held constant). In this sense, you can see that the regular-price demand has by far the largest effect on profit. The other two inputs, maximum supply and sale-price demand, have much smaller effects. Identifying important input variables is important for real applications. If a random input has a large effect on an important output, then it is probably worth the time and money to learn more about this input and possibly reduce the amount of uncertainty involving it. ■

PROBLEMS

Level A

16. If you add several normally distributed random numbers, the result is normally distributed, where the mean of the sum is the sum of the individual means, and the variance of the sum is the sum of the individual variances. (Remember that variance is the square of standard deviation.) This is a difficult result to prove mathematically, but it is easy to demonstrate with simulation. To do so, run a simulation where you add three normally distributed random numbers, each with mean 100 and standard deviation 10. Your single output variable should be the sum of these three numbers. Verify with @RISK that the distribution of this output is approximately normal with mean 300 and variance 300 (hence, standard deviation $\sqrt{300} = 17.32$).

17. In Problem 11 from the previous section, we stated that the damage amount is normally distributed. Suppose instead that the damage amount is triangularly distributed with parameters 500, 1500, and 7000. That is, the damage in an accident can be as low as $500 or as high as $7000, the most likely value is $1500, and there is definite skewness to the right. (It turns out, as you can verify in @RISK, that the mean of this distribution is $3000, the same as in Problem 11.) Use @RISK to simulate the amount you pay for damage. Run 5000 iterations. Then answer the following questions. In each case, explain how the indicated event would occur.

a. What is the probability that you pay a positive amount but less than $750?

b. What is the probability that you pay more than $600?

c. What is the probability that you pay exactly $1000 (the deductible)?

18. Continuing the previous problem, assume, as in Problem 11, that the damage amount is *normally* distributed with mean $3000 and standard deviation $750. Run @RISK with 5000 iterations to simulate the amount you pay for damage. Compare your results with those in the previous problem. Does it appear to matter whether you assume a triangular distribution or a normal distribution for damage amounts? Why isn't this a totally fair comparison? (*Hint*: Use @RISK's Define Distributions tool to find the standard deviation for the triangular distribution.)

19. In Problem 12 of the previous section, suppose that the demand for cars is normally distributed with mean 100 and standard deviation 15. Use @RISK to determine the "best" order quantity—in this case, the one with the largest mean profit. Using the statistics and/or graphs from @RISK, discuss whether this order quantity would be considered best by the car dealer. (The point is that a decision maker can use more than just *mean* profit in making a decision.)

20. Use @RISK to analyze the sweatshirt situation in Problem 14 of the previous section. Do this for the discrete distributions given in the problem. Then do it for normal distributions. For the normal case, assume that the regular demand is normally distributed with mean 9800 and standard deviation 1300 and that the demand at the reduced price is normally distributed with mean 3800 and standard deviation 1400.

Level B

21. Although the normal distribution is a reasonable input distribution in many situations, it does have two potential drawbacks: (1) it allows negative values, even though they may be extremely improbable, and (2) it is a symmetric distribution. Many situations are modeled better with a distribution that allows only positive values and is skewed to the right. Two of these that have been used in many real applications are the gamma and lognormal distributions. @RISK enables you to generate observations from each of these distributions. The @RISK function for the gamma distribution is RISKGAMMA, and it takes two arguments, as in **=RISKGAMMA(3,10)**. The first argument, which must be positive, determines the shape. The smaller it is, the more skewed the distribution is to the right; the larger it is, the more symmetric the distribution is. The second argument determines the scale, in the sense that the product of it and the first argument equals the mean of the distribution. (The mean in this example is 30.) Also, the product of the second argument and the square root of the first argument is the standard deviation of the distribution. (In this example, it is $\sqrt{3}(10) = 17.32$.) The @RISK function for the lognormal distribution is RISKLOGNORM. It has two arguments, as in **=RISKLOGNORM(40,10)**. These arguments are the mean and standard deviation of the distribution. Rework Example 15.2 for the following demand distributions. Do the simulated outputs have any different qualitative properties with these skewed distributions than with the triangular distribution used in the example?

a. Gamma distribution with parameters 2 and 85

b. Gamma distribution with parameters 5 and 35

c. Lognormal distribution with mean 170 and standard deviation 60

15-6 THE EFFECTS OF INPUT DISTRIBUTIONS ON RESULTS

In Section 15-2, we discussed input distributions. The randomness in input variables causes the variability in the output variables. We now briefly explore whether the choice of input distribution(s) makes much difference in the distribution of an output variable such as profit. This is an important question. If the choice of input distributions doesn't matter much, then you do not need to agonize over this choice. However, if it *does* make a difference, then you have to be more careful about choosing an appropriate input distribution for any particular situation. Unfortunately, it is impossible to answer the question definitively. The best we can say in general is, "It depends." Some models are more sensitive to changes in the shape or parameters of input distributions than others. Still, the issue is worth exploring.

We discuss two types of sensitivity analysis in this section. First, we check whether the shape of the input distribution matters. In the Walton Bookstore example, we assumed a triangularly distributed demand with some skewness. Are the results basically the same if a symmetric distribution such as the normal distribution is used instead? Second, we check whether the *independence* of input variables that have been assumed implicitly to this point is crucial to the output results. Many random quantities in real situations are *not* independent; they are positively or negatively correlated. Fortunately, @RISK enables you to build correlation into a model. We analyze the effect of this correlation.

15-6a Effect of the Shape of the Input Distribution(s)

We first explore the effect of the shape of the input distribution(s). As Example 15.5 indicates, if parameters that allow for a fair comparison are used, the shape can have a relatively minor effect.

EXAMPLE 15.5 EFFECT OF DEMAND DISTRIBUTION AT WALTON BOOKSTORE

We continue to explore the demand for calendars at Walton Bookstore. We keep the same unit cost, unit price, and unit refund for leftovers as in Example 15.3. However, in that example we assumed a triangular distribution for demand with parameters 100, 175, and 300. Assuming that Walton orders 200 calendars, is the distribution of profit affected if a *normal* distribution of demand is used instead?

Objective To see whether a triangular distribution with some skewness gives the same profit distribution as a normal distribution for demand.

Where Do the Numbers Come From?

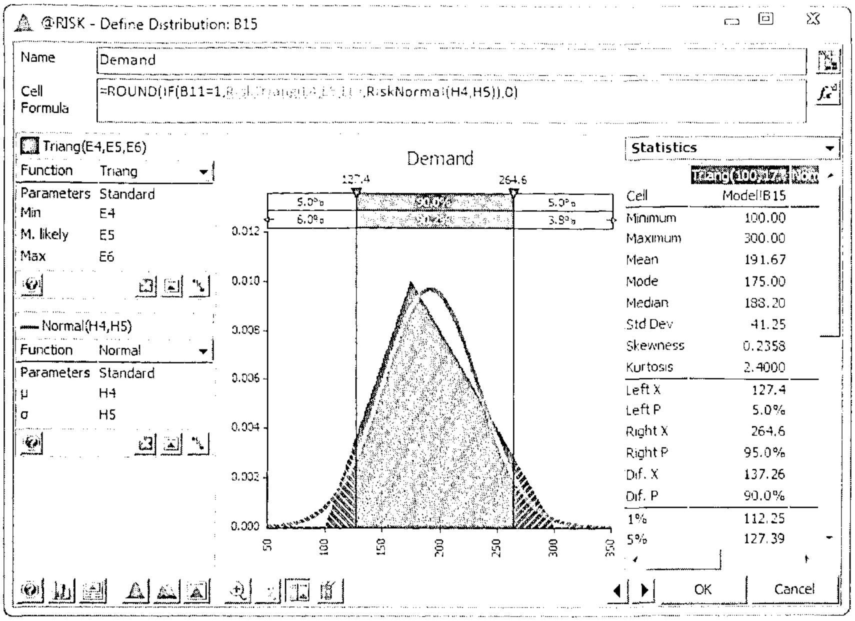
The numbers here are the same as in Example 15.3. However, as discussed next, the parameters of the normal distribution are chosen to provide a fair comparison with the triangular distribution used earlier.

Solution

For a fair comparison of alternative input distributions, the distributions should have (at least approximately) equal means and standard deviations.

It is important in this type of analysis to make a fair comparison. When you select a normal distribution for demand, you must choose a mean and standard deviation for this distribution. Which values should you choose? It seems only fair to choose the *same* mean and standard deviation that the triangular distribution has. To find the mean and standard deviation for a triangular distribution with given minimum, most likely, and maximum values, you can take advantage of @RISK's Define Distributions tool. Select any blank cell, click the Define Distributions button, select the triangular distribution, and enter the parameters 100, 175, and 300. You will see that the mean and standard deviation are 191.67 and 41.248, respectively. Therefore, for a fair comparison you should use a normal distribution with mean 191.67 and standard deviation 41.248. In fact, @RISK allows you to see a comparison of these two distributions, as in Figure 15.46. To get this chart, click the Add Overlay button, select the normal distribution from the gallery, and enter 191.67 and 41.248 as its mean and standard deviation.

Figure 15.46
Triangular and Normal Distributions for Demand



Developing the Simulation Model

The Walton Model with Alternative Input Distributions

The logic in this model is almost exactly the same as before. (See Figure 15.47 and the file **Ordering Calendars - Different Demand Distributions.xlsx**.) However, a clever use of the RISKSIMTABLE function allows you to run two simulations at once, one for the triangular distribution and one for the corresponding normal distribution. The following two steps are required.

Look for ways to use the RISKSIMTABLE function. It can really improve efficiency because it runs several simulations at once.

1. **RISKSIMTABLE function.** It is useful to index the two distributions as 1 and 2. To indicate that you want to run the simulation with both of them, enter the formula

   =RISKSIMTABLE({1,2})

   in cell B11. Note that when you enter actual numbers in this function, rather than cell references, you must put curly brackets around the list.

2. **Demand.** When the value in cell B11 is 1, the demand distribution is triangular. When it is 2, the distribution is normal. Therefore, enter the formula

   =ROUND(IF(B11=1,RISKTRIANG(E4,E5,E6),RISKNORMAL(H4,H5)),0)

   in cell B15. The effect is that the first simulation will use the triangular distribution, and the second will use the normal distribution.

Running the Simulation

The only @RISK setting to change is the number of simulations. It should now be set to 2, the number of values in the RISKSIMTABLE formula. Other than this, you run the simulation exactly as before.

Figure 15.47
@RISK Model for Comparing Two Input Distributions

	A	B	C	D	E	F	G	H	I
1	Simulation of Walton's Bookstore using @RISK - two possible demand distributions								
2									
3	Cost data			Demand distribution 1 - triangular			Demand distribution 2 - normal		
4	Unit cost	$7.50		Minimum	100		Mean	191.67	
5	Unit price	$10.00		Most likely	175		Stdev	41.248	
6	Unit refund	$2.50		Maximum	300				
7									
8	Decision variable								
9	Order quantity	200							
10									
11	Demand distribution to use	1	←	Formula is = RiskSimtable({1,2})					
12									
13	Simulated quantities								
14		Demand	Revenue	Cost	Refund	Profit			
15		209	$2,000	$1,500	$0	$500			
16									
17	Summary measures of profit from @RISK - based on 1000 iterations for each simulation								
18	Simulation	1	2						
19	Distribution	Triangular	Normal				Range names used:		
20	Minimum	–$235.00	–$632.50				Order_quantity	=Model!$B$9	
21	Maximum	$500.00	$500.00				Profit	=Model!$F$15	
22	Average	$337.51	$342.79				Unit_cost	=Model!$B$4	
23	Standard deviation	$189.06	$201.88				Unit_price	=Model!$B$5	
24	5th percentile	–$47.50	–$70.00				Unit_refund	=Model!$B$6	

Discussion of the Simulation Results

The comparison is shown numerically in Figure 15.48 and graphically in Figure 15.49. As you can see, there is more chance of really low profits when the demand distribution is normal, but each simulation results in the same maximum profit. Both of these statements make sense. The normal distribution, being unbounded on the left, allows for very low demands, and these occasional low demands result in very low profits. On the other side, Walton's maximum profit is $500 regardless of the input distribution (provided that it allows demands greater than the order quantity). This occurs when Walton's sells all it orders, in which case excess demand has no effect on profit. Note that the mean profits for the two distributions differ by only about $5.

Figure 15.48 Summary Results for Comparison Model

@RISK - Results Summary

Inputs Outputs

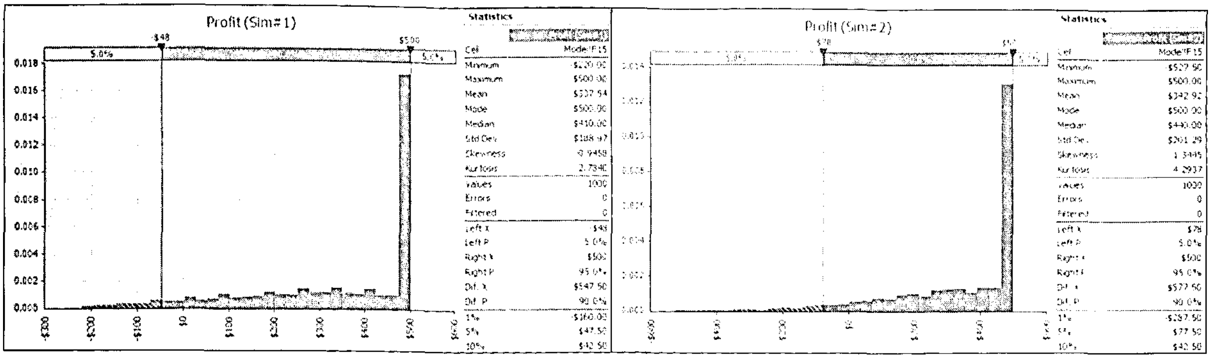
Simulation Results For Outputs: Inputs = 3, Outputs = 1, Iterations = 1000, Simulations = 2, Runtime= 00:00:01

Name	Cell	Sim#	Graph	Min	Mean	Max	Std Dev	5%	95%
[illegible]	[illegible]	[illegible]	-300 600	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Profit	F15	2	-800 600	-$632.50	$342.79	$500.00	$201.88	-$70.00	$500.00

Close

It is probably safe to conclude that the profit distribution in this model is not greatly affected by the choice of demand distribution, at least not when (1) the candidate input distributions have the same mean and standard deviation, and (2) their shapes are not *too* dissimilar. We would venture to guess that this general conclusion about insensitivity

Figure 15.49 Graphical Results for Comparison Model



of output distributions to shapes of input distributions can be made in many simulation models. However, it is always worth checking, as we have done here, especially when there is a lot of money at stake. ■

FUNDAMENTAL INSIGHT

Shape of the Output Distribution

Predicting the shape of the output distribution from the shape(s) of the input distribution(s) is difficult. For example, normally distributed inputs don't necessarily produce normally distributed outputs. It is also difficult to predict how sensitive the shape of the output distribution is to the shape(s) of the input distribution(s). For example, normally and triangularly distributed inputs (with the same means and standard deviations) are likely to lead to similar output distributions, but there could be differences, say, in the tails of the output distributions. In any case, you should examine the *entire* output distribution carefully, not just a few of its summary measures.

15-6b Effect of Correlated Input Variables

Input variables in real-world problems are often correlated, which makes the material in this section particularly important.

Until now, all of the random numbers generated with @RISK functions have been probabilistically independent. This means, for example, that if a random value in one cell is much larger than its mean, the random values in other cells are completely unaffected. They are no more likely to be abnormally large or small than if the first value had been average or below average. Sometimes, however, independence is unrealistic. In such cases, **correlated inputs** are more appropriate. If they are positively correlated, then large numbers will tend to go with large numbers, and small with small. If they are negatively correlated, then large will tend to go with small and small with large. As an example, you might expect daily stock price changes for two companies in the same industry to be positively correlated. If the price of one oil company increases, you might expect the price of another oil company to increase as well. You can create correlated inputs in @RISK with the **RISKCORRMAT** function, as we illustrate in the following continuation of the Walton example.

EXAMPLE **15.6 CORRELATED DEMANDS FOR TWO CALENDARS AT WALTON BOOKSTORE**

Suppose that Walton Bookstore must order two different calendars. To simplify the example, we assume that the calendars each have the same unit cost, unit selling price, and unit refund value as in previous examples. Also, we assume that each has a triangularly distributed demand with parameters 100, 175, and 300. However, we now assume they are "substitute" products, so that their demands are negatively correlated. This simply means that if a customer buys one, the customer is not likely to buy the other. Specifically, we assume a correlation of –0.9 between the two demands. How do these correlated inputs affect the distribution of profit, as compared to the situation where the demands are uncorrelated (correlation 0) or very *positively* correlated (correlation 0.9)?

Objective To see how @RISK enables us to simulate correlated demands, and to see the effect of correlated demands on profit.

Where Do the Numbers Come From?

The only new input here is the correlation. It is probably negative because the calendars are substitute products, but it is a difficult number to estimate accurately. This is a good candidate for a sensitivity analysis.

Solution

The key to building in correlation is @RISK's RISKCORRMAT (correlation matrix) function. To use this function, you must include a correlation matrix in the model, as shown in the range J5:K6 of Figure 15.50. (See the file **Ordering Calendars - Correlated Demands .xlsx.**) A correlation matrix must always have 1's along its diagonal (because a variable is always perfectly correlated with itself) and the correlations between variables elsewhere. Also, the matrix must be symmetric, so that the correlations above the diagonal are a mirror image of those below it. (You can enforce this by entering the formula **=J6** in cell K5. Alternatively, @RISK allows you to enter the correlations only below the diagonal, or only above the diagonal, and it then infers the mirror images.)

Figure 15.50 Simulation Model with Correlated Demands

	A	B	C	D	E	F	G	H	I	J	K
1	Simulation of Walton's Bookstore using @RISK - correlated demands										
2											
3	Cost data - same for each product			Demand distribution for each product- triangular					Correlation matrix between demands		
4	Unit cost	$7.50		Minimum	100					Product 1	Product 2
5	Unit price	$10.00		Most likely	175				Product 1	1	–0.9
6	Unit refund	$2.50		Maximum	300				Product 2	–0.9	1
7											
8	Decision variables					Note RISKSIMTABLE function in cell J6.			Possible correlations to try		
9	Order quantity 1	200							–0.9	0	0.9
10	Order quantity 2	200									
11									Range names used:		
12	Simulated quantities								Order_quantity_1	=Model!$B$9	
13		Demand	Revenue	Cost	Refund	Profit			Order_quantity_2	=Model!$B$10	
14	Product 1	136	$1,360	$1,500	$160	$20			Profit	=Model!$F$16	
15	Product 2	177	$1,770	$1,500	$58	$328			Unit_cost	=Model!$B$4	
16	Totals	313	$3,130	$3,000	$218	$348			Unit_price	=Model!$B$5	
17									Unit_refund	=Model!$B$6	
18	Summary measures of profit from @RISK - based on 1000 iterations										
19	Simulation	1	2	3							
20	Correlation	–0.9	0	0.9							
21	Minimum	$250.00	–$282.50	–$425.00							
22	Maximum	$1,000.00	$1,000.00	$1,000.00							
23	Average	$675.03	$675.03	$675.03							
24	Standard deviation	$159.27	$264.41	$365.90							
25	5th percentile	$392.50	$167.50	–$42.50							
26	95th percentile	$917.50	$1,000.00	$1,000.00							

The RISKCORRMAT function is "tacked on" as an extra argument to a typical random @RISK function.

To enter random values in any cells that are correlated, you start with a typical @RISK formula, such as

=RISKTRIANG(E4,E5,E6)

Then you add an extra argument, the RISKCORRMAT function, as follows:

=RISKTRIANG(E4,E5,E6,RISKCORRMAT(J5:K6,1))

The first argument of the RISKCORRMAT function is the correlation matrix range. The second is an index of the variable. In this example, the first calendar demand has index 1 and the second has index 2.

@RISK Function: RISKCORRMAT

This function enables you to correlate two or more input variables. The function has the form ***RISKCORRMAT(CorrMat,Index)****, where CorrMat is a matrix of correlations and* Index *is an index of the variable being correlated to others. For example, if there are three correlated variables,* Index *is 1 for the first variable, 2 is for the second, and 3 is for the third. The RISKCORRMAT function is not entered by itself. Rather, it is entered as the last argument of a random @RISK function, such as* ***=RISKTRIANG(10,15,30,RISKCORRMAT(CorrMat,2))****.*

The Walton Model with Correlated Demands

Developing the Simulation Model

Armed with this knowledge, the simulation model in Figure 15.50 is straightforward and can be developed as follows.

1. **Inputs.** Enter the inputs in the blue ranges in columns B and E.
2. **Correlation matrix.** For the correlation matrix in the range J5:H6, enter 1's on the diagonal, and enter the formula

   =J6

   in cell K5 (or leave cell K5 blank). Then enter the formula

   =RISKSIMTABLE(I9:K9)

   in cell J6. This allows you to simultaneously simulate negatively correlated demands, uncorrelated demands, and positively correlated demands.
3. **Order quantities.** Assume for now that the company orders the *same* number of each calendar, 200, so enter this value in cells B9 and B10. However, the simulation is set up so that you can experiment with any order quantities in these cells, including unequal values.
4. **Correlated demands.** Generate correlated demands by entering the formula

   Correlations in @RISK Models

   =ROUND(RISKTRIANG(E4,E5,E6,RISKCORRMAT(J5:K6,1)),0)

   in cell B14 for demand 1 and the formula

   =ROUND(RISKTRIANG(E4,E5,E6,RISKCORRMAT(J5:K6,2)),0)

   in cell B15 for demand 2. The only difference between these is the index of the variable being generated. The first has index 1; the second has index 2.
5. **Other formulas.** The other formulas in rows 14 and 15 are identical to ones developed in previous examples, so they aren't presented again here. The quantities in row 16 are simply sums of rows 14 and 15. Also, the only @RISK output we designated is the total profit in cell F16, but you can designate others as output cells if you like.

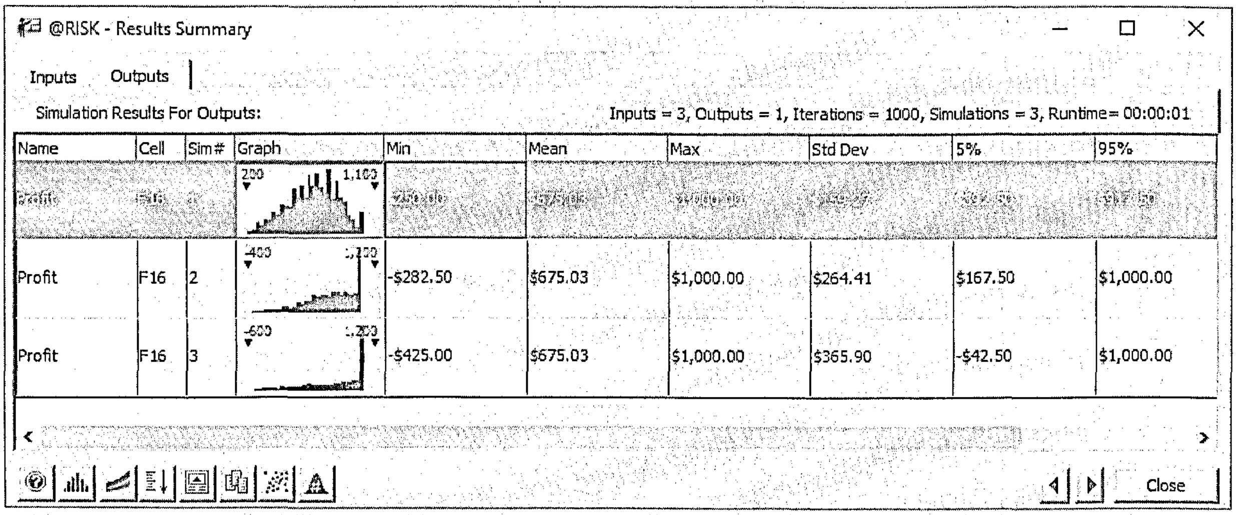
Running the Simulation

You should set up and run @RISK exactly as before. For this example, set the number of iterations to 1000 and the number of simulations to 3 (because three different correlations are being tested).

Discussion of the Simulation Results

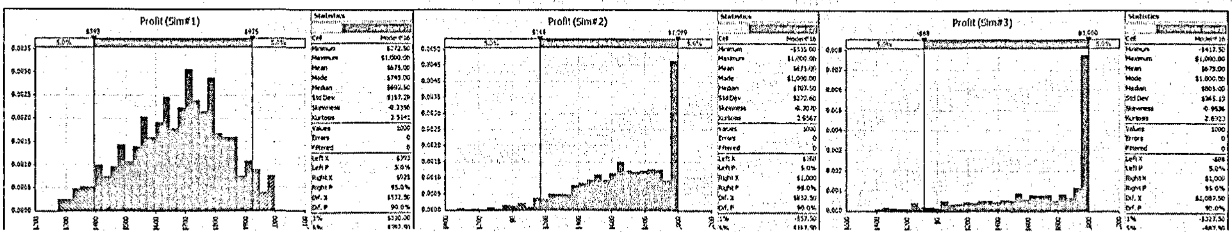
Selected numerical and graphical results are shown in Figures 15.51 and 15.52. You will probably be surprised to see that the *mean* total profit is the same, regardless of the correlation. This is no coincidence. In each of the three simulations, @RISK uses the *same* random numbers but "shuffles" them in different orders to get the correct correlations. This means that averages are unaffected. (The idea is that the average of the numbers 30, 26, and 48 is the same as the average of the numbers 48, 30, and 26.)

Figure 15.51 Summary Results for Correlated Model



© Cengage Learning

Figure 15.52 Graphical Results for Correlated Model



© Cengage Learning

However, the correlation has a definite effect on the *distribution* of total profit. You can see this in Figure 15.51, for example, where the standard deviation of total profit increases as the correlation goes from negative to zero to positive. This same increase in variability is apparent in the histograms in Figure 15.52. Do you see intuitively why this increase in variability occurs? It is basically the "Don't put all of your eggs in one basket" effect. When the correlation is negative, high demands for one product tend to cancel low demands for the other product, so extremes in profit are rare. However, when the correlation is positive, high demands for the two products tend to go together, as do low demands. These make extreme profits on either end much more likely.

This same phenomenon would occur if you simulated an investment portfolio containing two stocks. When the stocks are positively correlated, the portfolio is much riskier (more variability) than when they are negatively correlated. Of course, this is the reason for diversifying a portfolio.

Modeling Issues

With the RISKCORRMAT function, you can correlate random numbers from any distributions.

We illustrated the RISKCORRMAT function for triangularly distributed values. However, it can be used with any of @RISK's distributions by tacking on RISKCORRMAT as a last argument. You can even mix them. For example, assuming CMat is the range name for a 2×2 correlation matrix, you could enter the formulas

=RISKNORMAL(10,2,RISKCORRMAT(CMat,1))

and

=RISKUNIFORM(100,200,RISKCORRMAT(CMat,2))

into any two cells. When you run the simulation, @RISK generates a sequence of normally distributed random numbers based on the first formula and another sequence of uniformly distributed random numbers based on the second formula. Then it shuffles them in some complex way until their correlation is approximately equal to the specified correlation in the correlation matrix. ◘

FUNDAMENTAL INSIGHT

Correlated Inputs

When you enter random inputs in an @RISK simulation model and then run the simulation, each iteration generates *independent* values for the random inputs. If you know or suspect that some of the inputs are positively or negatively correlated, you should build this correlation structure into the model explicitly with the RISKCORRMAT function. This function might not change the mean of an output, but it can definitely affect the variability and shape of the output distribution.

PROBLEMS

Level A

22. Fizzy Company produces six-packs of soda cans. Each can is supposed to contain at least 12 ounces of soda. If the total weight in a six-pack is less than 72 ounces, Fizzy is fined $100 and receives no sales revenue for the six-pack. Each six-pack sells for $3.00. It costs Fizzy $0.02 per ounce of soda put in the cans. Fizzy can control the mean fill rate of its soda-filling machines. The amount put in each can by a machine is normally distributed with standard deviation 0.10 ounce.

a. Assume that the weight of each can in a six-pack has a 0.8 correlation with the weight of the other cans in the six-pack. What mean fill quantity maximizes expected profit per six-pack? Try mean fill rates from 12.00 to 12.35 in increments of 0.05.

b. If the weights of the cans in the six-pack are probabilistically independent, what mean fill quantity maximizes expected profit per six-pack? Try the same mean fill rates as in part **a.**

c. How can you explain the difference in the answers to parts **a** and **b**?

23. When you use @RISK's correlation feature to generate correlated random numbers, how can you verify that they are correlated? Try the following. Use the RISKCORRMAT function to generate two normally distributed random numbers, each with mean 100 and standard deviation 10, and with correlation 0.7. To run a simulation, you need an output variable, so sum these two numbers and designate the sum as an output variable. Now run @RISK with 500 iterations. Click @RISK's Excel Reports button and check the Simulation Data option to see the actual simulated data.

a. Use Excel's CORREL function to calculate the correlation between the two input variables. It should be close to 0.7. Then create a scatterplot of these two input variables. The plot should indicate a definite positive relationship.
b. Are the two input variables correlated with the output? Use Excel's CORREL function to find out. Interpret your results intuitively.

24. Work the previous problem, but make the correlation between the two inputs equal to −0.7. Explain how the results change.

25. Work Problem 23, but now make the second input variable triangularly distributed with parameters 50, 100, and 500. This time, verify not only that the correlation between the two inputs is approximately 0.7, but also that the shapes of the two input distributions are approximately what they should be: normal for the first and triangular for the second. Do this by creating histograms in Excel. The point is that you can use @RISK's RISKCORRMAT function to correlate random numbers from *different* distributions.

26. Suppose you are going to invest equal amounts in three stocks. The annual return from each stock is normally distributed with mean 0.01 (1%) and standard deviation 0.06. The annual return on your portfolio, the output variable of interest, is the average of the three stock returns. Run @RISK, using 1000 iterations, on each of the following scenarios.

a. The three stock returns are highly correlated. The correlation between each pair is 0.9.
b. The three stock returns are practically independent. The correlation between each pair is 0.1.
c. The first two stocks are moderately correlated. The correlation between their returns is 0.4. The third stock's return is negatively correlated with the other two. The correlation between its return and each of the first two is −0.8.
d. Compare the portfolio distributions from @RISK for these three scenarios. What do you conclude?
e. You might think of a fourth scenario, where the correlation between each *pair* of returns is a large negative number such as −0.8. But explain intuitively why this makes no sense. Try to run the simulation with these negative correlations and see what happens.

27. The effect of the shapes of input distributions on the distribution of an output can depend on the output function. For this problem, assume there are 10 input variables. The goal is to compare the case where these 10 inputs each have a normal distribution with mean 1000 and standard deviation 250 to the case where they each have a triangular distribution with parameters 600, 700, and 1700. (You can check with @RISK's Define Distributions window that even though this triangular distribution is very skewed, it has the same mean and approximately the same standard deviation as the normal distribution.) For each of the following outputs, run two @RISK simulations, one with the normally distributed inputs and one with the triangularly distributed inputs, and comment on the differences between the resulting output distributions. For each simulation run 1000 iterations.

a. Let the output be the *average* of the inputs.
b. Let the output be the *maximum* of the inputs.
c. Calculate the average of the inputs. Then the output is the minimum of the inputs if this average is less than 1000; otherwise, the output is the maximum of the inputs.

Level B

28. The Business School at State University currently has three parking lots, each containing 155 spaces. Two hundred faculty members have been assigned to each lot. On a peak day, an average of 70% of all lot 1 parking sticker holders show up, an average of 72% of all lot 2 parking sticker holders show up, and an average of 74% of all lot 3 parking sticker holders show up.

a. Given the current situation, estimate the probability that on a peak day, at least one faculty member with a sticker will be unable to find a spot. Assume that the number who show up at each lot is independent of the number who show up at the other two lots. Compare two situations: (1) each person can park only in the lot assigned to him or her, and (2) each person can park in any of the lots (pooling). (*Hint*: Use the RISKBINOMIAL function.)
b. Now suppose the numbers of people who show up at the three lots are highly correlated (correlation 0.9). How are the results different from those in part **a**?

15-7 CONCLUSION

Simulation has traditionally not received the attention it deserves in management science courses (or in business). The primary reason for this has been the lack of easy-to-use simulation software. Now, with Excel's built-in simulation capabilities, plus powerful and affordable add-ins such as @RISK, simulation is receiving its rightful emphasis. The world is full of uncertainty, which is what makes simulation so valuable. Simulation models provide important insights that are missing in models that do not incorporate uncertainty explicitly. In addition, simulation models are relatively easy to understand and develop. In this chapter we have illustrated the basic ideas of simulation, how to perform simulation with Excel built-in tools, and how @RISK greatly enhances Excel's basic capabilities. In the next chapter we will build on this knowledge to develop and analyze simulation models in a variety of business areas.

Summary of Key Terms

Term	Explanation	Excel	Page
Simulation model	Model with random inputs that affect one or more outputs, where the randomness is modeled explicitly		760
F9 key	The "recalc" key, used to make a spreadsheet recalculate		762
Probability distributions for input variables	Specification of the possible values and their probabilities for random input variables; these distributions must be specified in any simulation model		762
Uniform distribution	The flat distribution, where all values in a bounded continuum are equally likely		766
RAND function	Excel's built-in random number generator; generates uniformly distributed random numbers between 0 and 1	=RAND()	767
RANDBETWEEN function	Excel's built-in function for generating equally likely random integers over an indicated range	=RANDBETWEEN (*min,max*)	768
Freeze random numbers	Change "volatile" random numbers into "fixed" numbers	Copy range, paste it onto itself with the Paste Values option	770
@RISK random functions	A set of functions, including RISKNORMAL and RISKTRIANG, for generating random numbers from various distributions	=RISKNORMAL (*mean,stdev*) or =RISKTRIANG (*min,mostlikely,max*), for example	771
Discrete distribution	A general distribution where a discrete number of possible values and their probabilities are specified		772
Triangular distribution	Literally a triangle-shaped distribution, specified by a minimum value, a most likely value, and a maximum value		775
Replicating with Excel tools only	Useful when an add-in such as @RISK is not available	Develop simulation model, use a data table with any blank column input cell to replicate one or more outputs	791
@RISK	A powerful simulation add-in developed by Palisade	@RISK ribbon	794
RISKSIMTABLE function	Used to run an @RISK simulation model for several values of some variable, often a decision variable	=RISKSIMTABLE (*list*)	795

Term	Explanation	Excel	Page
RISKOUTPUT function	Used to indicate that a cell contains an output that will be tracked by @RISK	=RISKOUTPUT ("Profit") +Revenue-Cost, for example	798
Latin Hypercube sampling	An efficient way of simulating random numbers for a simulation model, where the results are more accurate than with other sampling methods		799
RISKCORRMAT function	Used to correlate two or more random input variables	=RISKNORMAL (100,10, RISKCORRMAT *(CorrMat,* 2)), for example	815
Correlated inputs	Random quantities, such as returns from stocks in the same industry, that tend to go together (or possibly go in opposite directions from one another)		815

PROBLEMS

Conceptual Questions

C.1. You are making several runs of a simulation model, each with a different value of some decision variable (such as the order quantity in the Walton calendar model), to see which decision value achieves the largest mean profit. Is it possible that one value beats another simply by random luck? What can you do to minimize the chance of a "better" value losing out to a "poorer" value?

C.2. If you want to replicate the results of a simulation model with Excel functions only, not @RISK, you can build a data table and let the column input cell be any blank cell. Explain why this works.

C.3. Suppose you simulate a gambling situation where you place many bets. On each bet, the distribution of your net winnings (loss if negative) is highly skewed to the left because there are some possibilities of really large losses but not much upside potential. Your only simulation output is the *average* of the results of all the bets. If you run @RISK with many iterations and look at the resulting histogram of this output, what will it look like? Why?

C.4. You plan to simulate a portfolio of investments over a multiyear period, so for each investment (which could be a particular stock or bond, for example), you need to simulate the change in its value for each of the years. How would you simulate these changes in a realistic way? Would you base it on historical data? What about correlations? Do you think the changes for different investments in a particular year would be correlated? Do you think changes for a particular investment in different years would be correlated? Do you think correlations would play a significant role in your simulation in terms of realism?

C.5. Big Hit Video must determine how many copies of a new video to purchase. Assume that the company's goal is to purchase a number of copies that maximizes its expected profit from the video during the next year. Describe how you would use simulation to shed light on this problem. Assume that each time a video is rented, it is rented for one day.

C.6. Many people who are involved in a small auto accident do not file a claim because they are afraid their insurance premiums will be raised. Suppose that City Farm Insurance has three rates. If you file a claim, you are moved to the next higher rate. How might you use simulation to determine whether a particular claim should be filed?

C.7. A building contains 1000 lightbulbs. Each bulb lasts at most five months. The company maintaining the building is trying to decide whether it is worthwhile to practice a "group replacement" policy. Under a group replacement policy, all bulbs are replaced every T months (where T is to be determined). Also, bulbs are replaced when they burn out. Assume that it costs \$0.05 to replace each bulb during a group replacement and \$0.20 to replace each burned-out bulb if it is replaced individually. How would you use simulation to determine whether a group replacement policy is worthwhile?

C.8. Why is the RISKCORRMAT function necessary? How does @RISK generate random inputs by default, that is, when RISKCORRMAT is not used?

C.9. Consider the claim that normally distributed inputs in a simulation model are bound to lead to normally distributed outputs. Do you agree or disagree with this claim? Defend your answer.

C.10. It is very possible that when you use a correlation matrix as input to the RISKCORRMAT function in an @RISK model, the program will inform you that this is an invalid correlation matrix. Provide an example of an obviously invalid correlation matrix involving at least three variables, and explain why it is invalid.

C.11. When you use a RISKSIMTABLE function for a decision variable, such as the order quantity in the Walton model, explain how this provides a "fair" comparison across the different values tested.

C.12. Consider a situation where there is a cost that is either incurred or not. It is incurred only if the value of some random input is less than a specified cutoff value. Why might a simulation of this situation give a very different average value of the cost incurred than a deterministic model that treats the random input as *fixed* at its mean? What does this have to do with the "flaw of averages"?

Level A

29. Six months before its annual convention, the American Medical Association must determine how many rooms to reserve. At this time, the AMA can reserve rooms at a cost of $150 per room. The AMA believes the number of doctors attending the convention will be normally distributed with a mean of 5000 and a standard deviation of 1000. If the number of people attending the convention exceeds the number of rooms reserved, extra rooms must be reserved at a cost of $250 per room.

a. Use simulation with @RISK to determine the number of rooms that should be reserved to minimize the expected cost to the AMA. Try possible values from 4100 to 4900 in increments of 100.

b. Redo part **a** for the case where the number attending has a triangular distribution with minimum value 2000, maximum value 7000, and most likely value 5000. Does this change the substantive results from part a?

30. You have made it to the final round of the show *Let's Make a Deal.* You know that there is a $1 million prize behind either door 1, door 2, or door 3. It is equally likely that the prize is behind any of the three doors. The two doors without a prize have nothing behind them. You randomly choose door 2. Before you see whether the prize is behind door 2, host Monty Hall opens a door that has no prize behind it. Specifically, suppose that before door 2 is opened, Monty reveals that there is no prize behind door 3. You now have the opportunity to switch and choose door 1. Should you switch? Simulate this situation 1000 times. For each replication use an @RISK function to generate the door that leads to the prize. Then use another @RISK function to generate the door that Monty will open. Assume that Monty plays as follows: Monty knows where the prize is and will open an empty door, but he cannot open door 2. If the prize is really behind door 2, Monty is equally likely to open door 1 or door 3. If the prize is really behind door 1, Monty must open door 3. If the prize is really behind door 3, Monty must open door 1.

31. A new edition of a very popular textbook will be published a year from now. The publisher currently has 2000 copies on hand and is deciding whether to do another printing before the new edition comes out. The publisher estimates that demand for the book during the next year is governed by the probability distribution in the file **P15_31.xlsx**. A production run incurs a fixed cost of $10,000 plus a variable cost of $15 per book printed. Books are sold for $130 per book. Any demand that cannot be met incurs a penalty cost of $20 per book, due to loss of goodwill. Up to 500 of any leftover books can be sold to Barnes & Noble for $35 per book. The publisher is interested in maximizing expected profit. The following print-run sizes are under consideration: 0 (no production run) to 16,000 in increments of 2000. What decision would you recommend? Use simulation with 1000 replications. For your optimal decision, the publisher can be 90% certain that the actual profit associated with remaining sales of the current edition will be between what two values?

32. A hardware company sells a lot of low-cost, high-volume products. For one such product, it is equally likely that annual unit sales will be low or high. If sales are low (60,000), the company can sell the product for $10 per unit. If sales are high (100,000), a competitor will enter and the company will be able to sell the product for only $8 per unit. The variable cost per unit has a 25% chance of being $6, a 50% chance of being $7.50, and a 25% chance of being $9. Annual fixed costs are $30,000.

a. Use simulation to estimate the company's expected annual profit.

b. Find a 95% interval for the company's annual profit, that is, an interval such that about 95% of the actual profits are inside it.

c. Now suppose that annual unit sales, variable cost, and unit price are equal to their respective *expected* values—that is, there is no uncertainty.

Determine the company's annual profit for this scenario.

d. Can you conclude from the results in parts **a** and **c** that the expected profit from a simulation is equal to the profit from the scenario where each input assumes its expected value? Explain.

33. W. L. Brown, a direct marketer of women's clothing, must determine how many telephone operators to schedule during each part of the day. W. L. Brown estimates that the number of phone calls received each hour of a typical eight-hour shift can be described by the probability distribution in the file **P15_33.xlsx**. Each operator can handle 15 calls per hour and costs the company \$20 per hour. Each phone call that is not handled is assumed to cost the company \$6 in lost profit. Considering the options of employing 6, 8, 10, 12, 14, or 16 operators, use simulation to determine the number of operators that minimizes the expected hourly cost (labor costs plus lost profits).

34. Assume that all of a company's job applicants must take a test, and that the scores on this test are normally distributed. The *selection ratio* is the cutoff point used by the company in its hiring process. For example, a selection ratio of 20% means that the company will accept applicants for jobs who rank in the top 20% of all applicants. If the company chooses a selection ratio of 20%, the average test score of those selected will be 1.40 standard deviations above average. Use simulation to verify this fact, proceeding as follows.

a. Show that if the company wants to accept only the top 20% of all applicants, it should accept applicants whose test scores are at least 0.842 standard deviation above average. (No simulation is required here. Just use the appropriate Excel normal function.)

b. Now generate 1000 test scores from a normal distribution with mean 0 and standard deviation 1. The average test score of those selected is the average of the scores that are at least 0.842. To determine this, use Excel's DAVERAGE function. To do so, put the heading Score in cell A3, generate the 1000 test scores in the range A4:A1003, and name the range A3:A1003 Data. In cells C3 and C4, enter the *labels* Score and >0.842. (The range C3:C4 is called the *criterion range*.) Then calculate the average of all applicants who will be hired by entering the formula **=DAVERAGE(Data, "Score", C3:C4)** in any cell. This average should be close to the theoretical average, 1.40. This formula works as follows. Excel finds all observations in the Data range that satisfy the criterion described in the range C3:C4 (Score>0.842). Then it averages the values in the Score column (the second argument of DAVERAGE) corresponding to these entries. See online help for more about Excel's database "D" functions.

c. What information would the company need to determine an optimal selection ratio? How could it determine the optimal selection ratio?

35. Lemington's is trying to determine how many Jean Hudson dresses to order for the spring season. Demand for the dresses is assumed to follow a normal distribution with mean 400 and standard deviation 100. The contract between Jean Hudson and Lemington's works as follows. At the beginning of the season, Lemington's reserves x units of capacity. Lemington's must take delivery for at least $0.8x$ dresses and can, if desired, take delivery on up to x dresses. Each dress sells for \$160 and Hudson charges \$50 per dress. If Lemington's does not take delivery on all x dresses, it owes Hudson a \$5 penalty for each unit of reserved capacity that is unused. For example, if Lemington's orders 450 dresses and demand is for 400 dresses, Lemington's will receive 400 dresses and owe Jean 400(\$50) + 50(\$5). How many units of capacity should Lemington's reserve to maximize its expected profit?

36. Dilbert's Department Store is trying to determine how many Hanson T-shirts to order. Currently the shirts are sold for \$21, but at later dates the shirts will be offered at a 10% discount, then a 20% discount, then a 40% discount, then a 50% discount, and finally a 60% discount. Demand at the full price of \$21 is believed to be normally distributed with mean 1800 and standard deviation 360. Demand at various discounts is assumed to be a multiple of full-price demand. These multiples, for discounts of 10%, 20%, 40%, 50%, and 60% are, respectively, 0.4, 0.7, 1.1, 2, and 50. For example, if full-price demand is 2500, then at a 10% discount customers would be willing to buy 1000 T-shirts. The unit cost of purchasing T-shirts depends on the number of T-shirts ordered, as shown in the file **P15_36.xlsx**. Use simulation to determine how many T-shirts the company should order. Model the problem so that the company first orders some quantity of T-shirts, then discounts deeper and deeper, as necessary, to sell all of the shirts.

Level B

37. The annual return on each of four stocks for each of the next five years is assumed to follow a normal distribution, with the mean and standard deviation for each stock, as well as the correlations between stocks, listed in the file **P15_37.xlsx**. You believe that the stock returns for these stocks in a given year are correlated, according to the correlation matrix given, but you believe the returns in different years are uncorrelated. For example, the returns for stocks 1 and 2 in year 1 have correlation 0.55, but the correlation between the return of stock 1 in year 1 and the return of stock 1 in year 2 is 0, and the correlation between the return of stock 1 in year 1 and the return of stock 2 in year 2 is also 0. The file has the formulas you might expect for this situation in the range C20:G23. You can check how the RISKCORRMAT function has been used in these formulas. Just so that there is an @RISK output cell, calculate the average of all returns in cell B25 and designate it as an @RISK output. (This cell is not really important for the problem, but it is included because @RISK requires at least one output cell.)

a. Using the model exactly as it stands, run @RISK with 1000 iterations. The question is whether the correlations in the simulated data are close to what they should be. To check this, go to @RISK's Report Settings and check the Input Data option before you run the simulation. This gives you all of the simulated returns on a new sheet. Then calculate correlations for all pairs of columns in the resulting Inputs Data Report sheet. (StatTools can be used to create a matrix of all correlations for the simulated data.) Comment on whether the correlations are different from what they should be.

b. Recognizing that this is a common situation (correlation within years, no correlation across years), @RISK allows you to model it by adding a *third* argument to the RISKCORRMAT function: the year index in row 19 of the **P15_37.xlsx** file. For example, the RISKCORRMAT part of the formula in cell C20 becomes **=RISKNORMAL($B5,$C5, RISKCORRMAT($B$12:$E$15,$B20,C$19))**. Make this change to the formulas in the range C20:G23, rerun the simulation, and redo the correlation analysis in part **a**. Verify that the correlations between inputs are now more in line with what they should be.

38. It is surprising (but true) that if 23 people are in the same room, there is about a 50% chance that at least two people will have the same birthday. Suppose you want to estimate the probability that if 30 people are in the same room, at least two of them will have the same birthday. You can proceed as follows.

a. Generate random birthdays for 30 different people. Ignoring the possibility of a leap year, each person has a 1/365 chance of having a given birthday (label the days of the year 1 to 365). You can use the RANDBETWEEN function to generate birthdays.

b. Once you have generated 30 people's birthdays, how can you tell whether at least two people have the same birthday? One way is to use Excel's RANK function. (You can learn how to use this function in Excel's online help.) This function returns the rank of a number relative to a given group of numbers. In the case of a tie, two numbers are given the same rank. For example, if the set of numbers is 4, 3, 2, 5, the RANK function returns 2, 3, 4, 1. (By default, RANK gives 1 to the *largest* number.) If the set of numbers is 4, 3, 2, 4, the RANK function returns 1, 3, 4, 1.

c. After using the RANK function, you should be able to determine whether at least two of the 30 people have the same birthday. What is the (estimated) probability that this occurs?

39. United Electric (UE) sells refrigerators for $400 with a one-year warranty. The warranty works as follows. If any part of the refrigerator fails during the first year after purchase, UE replaces the refrigerator for an average cost of $100. As soon as a replacement is made, another one-year warranty period begins for the customer. If a refrigerator fails outside the warranty period, we assume that the customer immediately purchases another UE refrigerator. Suppose that the amount of time a refrigerator lasts follows a normal distribution with a mean of 1.8 years and a standard deviation of 0.3 year.

a. Estimate the average profit per year UE earns from a customer.

b. How could the approach of this problem be used to determine the optimal warranty period?

40. A Flexible Savings Account (FSA) plan allows you to put money into an account at the beginning of the calendar year that can be used for medical expenses. This amount is not subject to federal tax. As you pay medical expenses during the year, you are reimbursed by the administrator of the FSA until the money is exhausted. From that point on, you must pay your medical expenses out of your own pocket. On the other hand, if you put more money into your FSA than the medical expenses you incur, this extra money is lost to you. Your annual salary is $80,000 and your federal income tax rate is 30%.

a. Assume that your medical expenses in a year are normally distributed with mean $2000 and standard deviation $500. Build an @RISK model in which the output is the amount of money left to you after paying taxes, putting money in an FSA, and paying any extra medical expenses. Experiment with the amount of money put in the FSA, using a RISKSIMTABLE function.

b. Rework part **a**, but this time assume a gamma distribution for your annual medical expenses. Use 16 and 125 as the two parameters of this distribution. These imply the same mean and standard deviation as in part **a**, but the distribution of medical expenses is now skewed to the right, which is probably more realistic. Using simulation, see whether you should now put more or less money in an FSA than in the symmetric case in part **a**.

41. At the beginning of each week, a machine is in one of four conditions: 1 = excellent; 2 = good; 3 = average;

4 = bad. The weekly revenue earned by a machine in state 1, 2, 3, or 4 is \$100, \$90, \$50, or \$10, respectively. After observing the condition of the machine at the beginning of the week, the company has the option, for a cost of \$200, of instantaneously replacing the machine with an excellent machine. The quality of the machine deteriorates over time, as shown in the file **P15_41.xlsx**. Four maintenance policies are under consideration:

- Policy 1: Never replace a machine.
- Policy 2: Immediately replace a bad machine.
- Policy 3: Immediately replace a bad or average machine.
- Policy 4: Immediately replace a bad, average, or good machine.

Simulate each of these policies for 50 weeks (using at least 250 iterations each) to determine the policy that maximizes expected weekly profit. Assume that the machine at the beginning of week 1 is excellent.

42. Simulation can be used to illustrate a number of results from statistics that are difficult to understand with nonsimulation arguments. One is the famous *central limit theorem*, which says that if you sample enough values from *any* population distribution and then average these values, the resulting average will be approximately normally distributed. Confirm this by using @RISK with the following population distributions (run a separate simulation for each): (a) discrete with possible values 1 and 2 and probabilities 0.2 and 0.8; (b) exponential with mean 1 (use the RISKEXPON function with the single argument 1); (c) triangular with minimum, most likely, and maximum values equal to 1, 9, and 10. (Note that each of these distributions is very skewed.) Run each simulation with 10 values in each average, and run 1000 iterations to simulate 1000 averages. Create a histogram of the averages to see whether it is indeed bell-shaped. Then repeat, using 30 values in each average. Are the histograms based on 10 values qualitatively different from those based on 30?

43. In statistics we often use observed data to test a hypothesis about a population or populations. The basic method uses the observed data to calculate a test statistic (a single number), as discussed in Chapter 9. If the magnitude of this test statistic is sufficiently large, the null hypothesis is rejected in favor of the research hypothesis. As an example, consider a researcher who believes teenage girls sleep longer than teenage boys on average. She collects observations on $n = 40$ randomly selected girls and $n = 40$ randomly selected boys. (Each observation is the average sleep time over several nights for a given person.) The averages are $\overline{X}_1 = 7.9$ hours for the girls and $\overline{X}_2 = 7.6$ hours for the boys. The standard deviation of the 40 observations for girls is $s_1 = 0.5$ hour; for the boys it is $s_2 = 0.7$ hour. The researcher, consulting Chapter 9, then calculates the test statistic

$$\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{s_1^2/40 + s_2^2/40}} = \frac{7.9 - 7.6}{\sqrt{0.25/40 + 0.49/40}} = 2.206$$

Based on the fact that 2.206 is "large," she claims that her research hypothesis is confirmed—girls do sleep longer than boys.

You are skeptical of this claim, so you check it out by running a simulation. In your simulation you assume that girls and boys have the *same* mean and standard deviation of sleep times in the entire population, say, 7.7 and 0.6. You also assume that the distribution of sleep times is normal. Then you repeatedly simulate observations of 40 girls and 40 boys from this distribution and calculate the test statistic. The question is whether the observed test statistic, 2.206, is "extreme." If it is larger than most or all of the test statistics you simulate, then the researcher is justified in her claim; otherwise, this large a statistic could have happened easily by chance, even if the girls and boys have identical population means. Use @RISK to see which of these possibilities occurs.

44. A technical note in the discussion of @RISK indicated that Latin Hypercube sampling is more efficient than Monte Carlo sampling. This problem allows you to see what this means. The file **P15_44.xlsx** gets you started. There is a single output cell, B5. You can enter any random value in this cell, such as **RISKNORMAL(500,100)**. There are already @RISK statistical formulas in rows 9–12 to calculate summary measures of the output for each of 10 simulations. On the @RISK ribbon, click on the button to the left of the "dice" button to bring up the Simulation Settings dialog box, click on the Sampling tab, and make sure the Sampling Type is Latin Hypercube. Run 10 simulations with at least 1000 iterations each, and then paste the results in rows 9–12 as *values* in rows 17–20. Next, get back in Simulations Settings and change the Sampling Type to Monte Carlo, run the 10 simulations again, and paste the results in rows 9–12 as values into rows 23–26. For each row, 17–20 and 23–26, summarize the 10 numbers in that row with AVERAGE and STDEV. What do you find? Why do we say that Latin Hypercube sampling is more efficient? (Thanks to Harvey Wagner at University of North Carolina for suggesting this problem.)

45. We are continually hearing reports on the nightly news about natural disasters—droughts in Texas, hurricanes in Florida, floods in California, and so on. We often hear that one of these was the "worst in over 30 years," or some such statement. Are natural disasters getting worse these days, or does it just appear so? How might you use simulation to answer this question? Here is one possible approach. Imagine that there are N areas of the country (or the world) that tend to have, to some extent, various types of weather phenomena each year. For example, hurricanes are always a potential problem for Florida, and fires are always a potential problem in southern California. You might model the severity of the problem for any area in any year by a normally distributed random number with mean 0 and standard deviation 1, where negative values are interpreted as good years and positive values are interpreted as bad years. (We suggest the normal distribution, but there is no reason other distributions couldn't be used instead.) Then you could simulate such values for all areas over a period of several years and keep track, say, of whether any of the areas have worse conditions in the current year than they have had in the past several years, where "several" could be 10, 20, 30, or any other number of years you want to test. What might you keep track of? How might you interpret your results?

CASE 15.1 Ski Jacket Production

Egress, Inc., is a small company that designs, produces, and sells ski jackets and other coats. The creative design team has labored for weeks over its new design for the coming winter season. It is now time to decide how many ski jackets to produce in this production run. Because of the lead times involved, no other production runs will be possible during the season. Predicting ski jacket sales months in advance of the selling season can be quite tricky. Egress has been in operation for only three years, and its ski jacket designs were quite successful in two of those years. Based on realized sales from the last three years, current economic conditions, and professional judgment, 12 Egress employees have independently estimated demand for their new design for the upcoming season. Their estimates are listed in Table 15.2.

Table 15.2 Estimated Demands

14,000	16,000
13,000	8000
14,000	5000
14,000	11,000
15,500	8000
10,500	15,000

To assist in the decision on the number of units for the production run, management has gathered the data in Table 15.3. Note that S is the price Egress charges retailers. Any ski jackets that do not sell during the season can be sold by Egress to discounters for V per jacket. The fixed cost of plant and equipment is F. This cost is incurred regardless of the size of the production run.

Table 15.3 Monetary Values

Variable production cost per unit (C):	$80
Selling price per unit (S):	$100
Salvage value per unit (V):	$30
Fixed production cost (F):	$100,000

Questions

1. Egress management believes that a normal distribution is a reasonable model for the unknown demand in the coming year. What mean and standard deviation should Egress use for the demand distribution?
2. Use a spreadsheet model to simulate 1000 possible outcomes for demand in the coming year. Based on these scenarios, what is the expected profit if Egress produces $Q = 7800$ ski jackets? What is the expected profit if Egress produces $Q = 12,000$ ski jackets? What is the standard deviation of profit in these two cases?
3. Based on the same 1000 scenarios, how many ski jackets should Egress produce to maximize expected profit? Call this quantity Q.
4. Should Q equal mean demand or not? Explain.
5. Create a histogram of profit at the production level Q. Create a histogram of profit when the production level Q equals mean demand. What is the probability of a loss greater than $100,000 in each case?

CASE 15.2 EBONY BATH SOAP

Management of Ebony, a leading manufacturer of bath soap, is trying to control its inventory costs. The weekly cost of holding one unit of soap in inventory is \$30 (one unit is 1000 cases of soap). The marketing department estimates that weekly demand averages 120 units, with a standard deviation of 15 units, and is reasonably well modeled by a normal distribution. If demand exceeds the amount of soap on hand, those sales are *lost*—that is, there is no backlogging of demand. The production department can produce at one of three levels: 110, 120, or 130 units per week. The cost of changing the production level from one week to the next is \$3000.

Management would like to evaluate the following production policy. If the current inventory is less than $L = 30$ units, they will produce 130 units in the next week. If the current inventory is greater than $U = 80$ units, they will produce 110 units in the next week. Otherwise, Ebony will continue at the previous week's production level.

Ebony currently has 60 units of inventory on hand. Last week's production level was 120.

Questions

1. Develop a simulation model for 52 weeks of operation at Ebony. Graph the inventory of soap over time. What is the total cost (inventory cost plus production change cost) for the 52 weeks?
2. Run the simulation for 500 iterations to estimate the average 52-week cost with values of U ranging from 30 to 80 in increments of 10. Keep $L = 30$ throughout.
3. Report the sample mean and standard deviation of the 52-week cost under each policy. Using the simulated results, is it possible to construct *valid* 90% confidence intervals for the average 52-week cost for each value of U? In any case, graph the average 52-week cost versus U. What is the best value of U for $L = 30$?
4. What other production policies might be useful to investigate?