12.0 INTRODUCTION

Chapter 1 discussed how the calculations in a spreadsheet can be viewed as a mathematical model that defines a functional relationship between various input variables (or independent variables) and one or more bottom-line performance measures (or dependent variables). The following equation expresses this relationship:

\[ Y = f(X_1, X_2, ..., X_k) \]

In many spreadsheets, the values of various input cells are determined by the person using the spreadsheet. These input cells correspond to the independent variables \( X_1, X_2, ..., X_k \) in the previous equation. Various formulas (represented by \( f(\) above) are entered in other cells of the spreadsheet to transform the values of the input cells into some bottom-line output (denoted by \( Y \) above). Simulation is a technique that is helpful in analyzing models in which the value to be assumed by one or more independent variables is uncertain.

This chapter discusses how to perform simulation using a popular commercial spreadsheet add-in called Risk Solver Platform, created and distributed by FrontLine Systems. A limited-life (15-day) trial version of Risk Solver Platform and related products may be downloaded from FrontLine's web site at www.solver.com. A non-expiring educational version of Risk Solver Platform is also available to students.

12.1 RANDOM VARIABLES AND RISK

In order to compute a value for the bottom-line performance measure of a spreadsheet model, each input cell must be assigned a specific value so that all the related calculations can be performed. However, some uncertainty often exists regarding the value that should be assumed by one or more independent variables (or input cells) in the spreadsheet. This is particularly true in spreadsheet models that represent future conditions. A random variable is any variable whose value cannot be predicted or set with certainty. Thus, many input variables in a spreadsheet model represent random variables whose actual values cannot be predicted with certainty.

For example, projections of the cost of raw materials, future interest rates, future numbers of employees, and expected product demand are random variables because their true values are unknown and will be determined in the future. If we cannot say with certainty what value one or more input variables in a model will assume, we also cannot say with certainty what value the dependent variable will assume. This uncertainty associated with the value of the dependent variable introduces an element of risk to the decision-making problem. Specifically, if the dependent variable represents some bottom-line performance measure that managers use to make decisions, and its value is uncertain, any decisions made on the basis of this value are based on uncertain (or incomplete) information. When such a decision is
made, some chance exists that the decision will not produce the intended results. This chance, or uncertainty, represents an element of risk in the decision-making problem.

The term “risk” also implies the potential for loss. The fact that a decision’s outcome is uncertain does not mean that the decision is particularly risky. For example, whenever we put money into a soft drink machine, there is a chance the machine will take our money and not deliver the product. However, most of us would not consider this risk to be particularly great. From past experience, we know that the chance of not receiving the product is small. But even if the machine takes our money and does not deliver the product, most of us would not consider this to be a tremendous loss. Thus, the amount of risk involved in a given decision-making situation is a function of the uncertainty in the outcome of the decision and the magnitude of the potential loss. A proper assessment of the risk present in a decision-making situation should address both of these issues, as the examples in this chapter will demonstrate.

12.2 WHY ANALYZE RISK?

Many spreadsheets built by business people contain estimated values for the uncertain input variables in their models. If a manager cannot say with certainty what value a particular cell in a spreadsheet will assume, this cell most likely represents a random variable. Ordinarily, the manager will attempt to make an informed guess about the values such cells will assume. The manager hopes that inserting the expected, or most likely, values for all the uncertain cells in a spreadsheet will provide the most likely value for the cell containing the bottom-line performance measure (Y). The problem with this type of analysis is that it tells the decision maker nothing about the variability of the performance measure.

For example, in analyzing a particular investment opportunity, we might determine that the expected return on a $1,000 investment is $10,000 within two years. But how much variability exists in the possible outcomes? If all the potential outcomes are scattered closely around $10,000 (say from $9,000 to $11,000), then the investment opportunity might still be attractive. If, on the other hand, the potential outcomes are scattered widely around $10,000 (say from –$30,000 up to +$50,000), then the investment opportunity might be unattractive. Although these two scenarios might have the same expected or average value, the risks involved are quite different. Thus, even if we can determine the expected outcome of a decision using a spreadsheet, it is just as important, if not more so, to consider the risk involved in the decision.

12.3 METHODS OF RISK ANALYSIS

Several techniques are available to help managers analyze risk. Three of the most common are best-case/worst-case analysis, what-if analysis, and simulation. Of these methods, simulation is the most powerful and, therefore, is the technique that we will focus on in this chapter. Although the other techniques might not be completely effective in risk analysis, they are probably used more often than simulation by most managers in business today. This is largely due to the fact that most managers are unaware of the spreadsheet’s ability to perform simulation and of the benefits provided by this technique. So before discussing simulation, let’s first briefly look at the other methods of risk analysis to understand their strengths and weaknesses.

12.3.1 Best-Case/Worst-Case Analysis

If we don’t know what value a particular cell in a spreadsheet will assume, we could enter a number that we think is the most likely value for the uncertain cell. If we enter such numbers for all the uncertain cells in the spreadsheet, we can easily calculate the most likely value of the bottom-line performance measure. This is also called the base-case scenario. However, this scenario gives us no information about how far away the actual outcome might be from this expected, or most likely, value.

One simple solution to this problem is to calculate the value of the bottom-line performance measure using the best-case, or most optimistic, and worst-case, or most pessimistic, values for the uncertain input cells. These additional scenarios show the range of possible values that might be assumed by the
bottom-line performance measure. As indicated in the earlier example about the $1,000 investment, knowing the range of possible outcomes is very helpful in assessing the risk involved in different alternatives. However, simply knowing the best-case and worst-case outcomes tells us nothing about the distribution of possible values within this range, nor does it tell us the probability of either scenario occurring.

Figure 12.1 displays several probability distributions that might be associated with the value of a bottom-line performance measure within a given range. Each of these distributions describes variables that have identical ranges and similar average values. But each distribution is very different in terms of the risk it represents to the decision maker. The appeal of best-case/worst-case analysis is that it is easy to do. Its weakness is that it tells us nothing about the shape of the distribution associated with the bottom-line performance measure. As we will see later, knowing the shape of the distribution of the bottom-line performance measure can be critically important in helping us answer a number of managerial questions.

![Possible distributions of performance measure values within a given range.](image)

**Figure 12.1**

### 12.3.2 What-If Analysis

Prior to the introduction of electronic spreadsheets in the early 1980s, the use of best-case/worst-case analysis was often the only feasible way for a manager to analyze the risk associated with a decision. This process was extremely time-consuming, error prone, and tedious, using only a piece of paper, pencil, and calculator to recalculate the performance measure of a model using different values for the uncertain inputs. The arrival of personal computers and electronic spreadsheets made it much easier for a manager to play out a large number of scenarios in addition to the best and worst cases—which is the essence of what-if analysis.

In what-if analysis, a manager changes the values of the uncertain input variables to see what happens to the bottom-line performance measure. By making a series of such changes, a manager can gain some insight into how sensitive the performance measure is to changes to the input variables. Although many managers perform this type of manual what-if analysis, it has three major flaws.
Chapter 12 Introduction to Simulation Using Risk Solver Platform

First, if the values selected for the independent variables are based only on the manager’s judgment, the resulting sample values of the performance measure are likely to be biased. That is, if several uncertain variables can each assume some range of values, it would be difficult to ensure that the manager tests a fair, or representative, sample of all possible combinations of these values. To select values for the uncertain variables that correctly reflect their random variations, the values must be randomly selected from a distribution, or pool, of values that reflects the appropriate range of possible values, as well as the appropriate relative frequencies of these variables.

Second, hundreds or thousands of what-if scenarios might be required to create a valid representation of the underlying variability in the bottom-line performance measure. No one would want to perform these scenarios manually nor would anyone be able to make sense of the resulting stream of numbers that would flash on the screen.

The third problem with what-if analysis is that the insight the manager might gain from playing out various scenarios is of little value when recommending a decision to top management. What-if analysis simply does not supply the manager with the tangible evidence (facts and figures) needed to justify why a given decision was made or recommended. Additionally, what-if analysis does not address the problem identified in our earlier discussion of best-case/worst-case analysis—it does not allow us to estimate the distribution of the performance measure in a formal enough manner. Thus, what-if analysis is a step in the right direction, but it is not quite a large enough step to allow managers to analyze risk effectively in the decisions they face.

12.3.3 Simulation

Simulation is a technique that measures and describes various characteristics of the bottom-line performance measure of a model when one or more values for the independent variables are uncertain. If any independent variables in a model are random variables, the dependent variable (Y) also represents a random variable. The objective in simulation is to describe the distribution and characteristics of the possible values of the bottom-line performance measure Y, given the possible values and behavior of the independent variables $X_1, X_2, ..., X_k$.

The idea behind simulation is similar to the notion of playing out many what-if scenarios. The difference is that the process of assigning values to the cells in the spreadsheet that represent random variables is automated so that: (1) the values are assigned in a non-biased way, and (2) the spreadsheet user is relieved of the burden of determining these values. With simulation, we repeatedly and randomly generate sample values for each uncertain input variable ($X_1, X_2, ..., X_k$) in our model and then compute the resulting value of our bottom-line performance measure (Y). We can then use the sample values of Y to estimate the true distribution and other characteristics of the performance measure Y. For example, we can use the sample observations to construct a frequency distribution of the performance measure, to estimate the range of values over which the performance measure might vary, to estimate its mean and variance, and to estimate the probability that the actual value of the performance measure will be greater than (or less than) a particular value. All these measures provide greater insight into the risk associated with a given decision than a single value calculated based on the expected values for the uncertain independent variables.

On Uncertainty and Decision-Making...

"Uncertainty is the most difficult thing about decision-making. In the face of uncertainty, some people react with paralysis, or they do exhaustive research to avoid making a decision. The best decision-making happens when the mental environment is focused. In a physical environment, you focus on something physical. In tennis, that might be the spinning seams of the ball. In a mental environment, you focus on the facts at hand. That fine-tuned focus doesn’t leave room for fears and doubts to enter. Doubts knock at the door of our consciousness, but you don’t have to have them in for tea and crumpets." -- Timothy Gallwey, author of The Inner Game of Tennis and The Inner Game of Work.
12.4 A CORPORATE HEALTH INSURANCE EXAMPLE

The following example demonstrates the mechanics of preparing a spreadsheet model for risk analysis using simulation. The example presents a fairly simple model to illustrate the process and give a sense of the amount of effort involved. However, the process for performing simulation is basically the same regardless of the size of the model.

Lisa Pon has just been hired as an analyst in the corporate planning department of Hungry Dawg Restaurants. Her first assignment is to determine how much money the company needs to accrue in the coming year to pay for its employees’ health insurance claims. Hungry Dawg is a large, growing chain of restaurants that specializes in traditional southern foods. The company has become large enough that it no longer buys insurance from a private insurance company. The company is now self-insured, meaning that it pays health insurance claims with its own money (although it contracts with an outside company to handle the administrative details of processing claims and writing checks).

The money the company uses to pay claims comes from two sources: employee contributions (or premiums deducted from employees’ paychecks), and company funds (the company must pay whatever costs are not covered by employee contributions). Each employee covered by the health plan contributes $125 per month. However, the number of employees covered by the plan changes from month to month as employees are hired and fired, quit, or simply add or drop health insurance coverage. A total of 18,533 employees were covered by the plan last month. The average monthly health claim per covered employee was $250 last month.

An example of how most analysts would model this problem is shown in Figure 12.2 on the following page (and in the file Fig12-2.xls). The spreadsheet begins with a listing of the initial conditions and assumptions for the problem. For example, cell D5 indicates that 18,533 employees are currently covered by the health plan, and cell D6 indicates that the average monthly claim per covered employee is $250. The average monthly contribution per employee is $125, as shown in cell D7. The values in cells D5 and D6 are unlikely to stay the same for the entire year. Thus, we need to make some assumptions about the rate at which these values are likely to increase during the year. For example, we might assume that the number of covered employees will increase by about 2% per month, and that the average claim per employee will increase at a rate of 1% per month. These assumptions are reflected in cells F5 and F6. The average contribution per employee is assumed to be constant over the coming year.

Using the assumed rate of increase in the number of covered employees (cell F5), we can create formulas for cells B11 through B22 that cause the number of covered employees to increase by the assumed amount each month. (The details of these formulas are covered later.) The expected monthly employee contributions shown in column C are calculated as $125 times the number of employees in each month. We can use the assumed rate of increase in average monthly claims (cell F6) to create formulas for cells D11 through D22 that cause the average claim per employee to increase at the assumed rate. The total claims for each month (shown in column E) are calculated as the average claim figures in column D times the number of employees for each month in column B. Because the company must pay for any claims that are not covered by the employee contributions, the company cost figures in column G are calculated as the total claims minus the employee contributions (column E minus column C). Finally, cell G23 sums the company cost figures listed in column G, and shows that the company can expect to contribute $36,125,850 of its revenues toward paying the health insurance claims of its employees in the coming year.

12.4.1 A Critique of the Base Case Model

Now, let’s consider the model we just described. The example model assumes that the number of covered employees will increase by exactly 2% each month and that the average claim per covered employee will increase by exactly 1% each month. Although these values might be reasonable approximations of what might happen, they are unlikely to reflect exactly what will happen. In fact, the number of employees
covered by the health plan each month is likely to vary randomly around the average increase per month— that is, the number might decrease in some months and increase by more than 2% in others. Similarly, the average claim per covered employee might be lower than expected in certain months and higher than expected in others.

Both of these figures are likely to exhibit some uncertainty or random behavior, even if they do move in the general upward direction assumed throughout the year. So, we cannot say with certainty that the total cost figure of $36,125,850 is exactly what the company will have to contribute toward health claims in the coming year. It is simply a prediction of what might happen. The actual outcome could be smaller or larger than this estimate. Using the original model, we have no idea how much larger or smaller the actual result could be—nor do we have any idea of how the actual values are distributed around this estimate. We do not know if there is a 10%, 50%, or 90% chance of the actual total costs exceeding this estimate. To determine the variability or risk inherent in the bottom-line performance measure of total company costs, we will apply the technique of simulation to our model.

![Original corporate health insurance model with expected values for uncertain variables.](image)

**Figure 12.2**

### 12.5 SPREADSHEET SIMULATION USING Risk Solver Platform

To perform simulation in a spreadsheet, we must first place a **random number generator** (RNG) formula in each cell that represents a random, or uncertain, independent variable. Each RNG provides a sample observation from an appropriate distribution that represents the range and frequency of possible values for the variable. Once the RNGs are in place, new sample values are provided automatically each time the spreadsheet is recalculated. We can recalculate the spreadsheet *n* times, where *n* is the desired number of replications or scenarios, and the value of the bottom-line performance measure will be stored.
after each replication. We can analyze these stored observations to gain insights into the behavior and characteristics of the performance measure.

The process of simulation involves a lot of work but, fortunately, the spreadsheet can do most of the work for us fairly easily. In particular, the spreadsheet add-in package Risk Solver Platform is designed specifically to make spreadsheet simulation a simple process. Risk Solver Platform provides the following capabilities, which are not otherwise available while working in Excel: additional functions that are helpful in generating the random numbers needed in simulation; additional commands that are helpful in setting up and running the simulation; and graphical and statistical summaries of the simulation data. As we shall see, these capabilities make simulation a relatively easy technique to apply in spreadsheets.

### 12.5.1 Starting Risk Solver Platform

If you are running Risk Solver Platform from a local area network (LAN) or in a computer lab, your instructor or LAN coordinator should give you directions on how to access this software. If you have installed Risk Solver Platform on your own computer, the Risk Solver Platform tab should automatically appear on the Ribbon as shown in Figure 12.2. You can also load (or unload) Risk Solver Platform manually from within Excel as follows:

1. Click the Office button.
2. Click Excel Options.
3. Select Add-Ins.
4. Select (or unselect) the Risk Solver Platform COM add-in and also the Risk Solver Platform Excel Add-In.
5. Click OK.

The Risk Solver Platform tab contains the custom ribbon shown in Figure 12.2. We will refer to the various icons on this ribbon throughout this chapter.

### 12.6 RANDOM NUMBER GENERATORS

As mentioned earlier, the first step in spreadsheet simulation is to place an RNG formula in each cell that contains an uncertain value. Each of these formulas will generate (or return) a number that represents a randomly selected value from a distribution, or pool, of values. The distributions that these samples are taken from should be representative of the underlying pool of values expected to occur in each uncertain cell.

Risk Solver Platform provides several “Psi” functions that can be used to create the RNGs required for simulating a model. (The “Psi” prefix on these functions stands for Polymorphic Spreadsheet Interpreter, which is the technology FrontLine Systems developed and uses in Risk Solver Platform to recalculate Excel workbooks extremely quickly.) Figure 12.3 describes some of the most common RNGs. These functions allow us to generate a variety of random numbers easily. For example, if we think that the behavior of an uncertain cell could be modeled as a normally distributed random variable with a mean of 125 and standard deviation of 10, then according to Figure 12.3 we could enter the formula =PsiNormal(125,10) in this cell. (The arguments in this function could also be formulas and could refer to other cells in the spreadsheet.) After this formula is entered, Risk Solver Platform will randomly generate or select value from a normal distribution with a mean of 125 and standard deviation of 10 for this cell whenever the spreadsheet is recalculated.

Similarly, a cell in our spreadsheet might have a 30% chance of assuming the value 10, a 50% chance of assuming the value 20, and a 20% chance of assuming the value 30. As noted in Figure 12.3, we could use the formula =PsiDiscrete({10,20,30},{0.3,0.5,0.2}) to model the behavior of this random variable. If we recalculated the spreadsheet many times, this formula would return the value 10 approximately 30% of the time, the value 20 approximately 50% of the time, and the value 30 approximately 20% of the time.

The arguments, or parameters, required by the RNG functions allow us to generate random numbers from distributions with a wide variety of shapes. Figures 12.4 and 12.5 illustrate some example
distributions. Additional information about these and other RNGs provided by Risk Solver Platform is available in the Risk Solver Platform user manual and in the on-line Help facility in Risk Solver Platform.

**Software Note**

A listing of all the available Risk Solver Platform RNG functions is available in Excel. To view this listing:
1. Select an empty cell in a worksheet.
2. Click Formulas, Insert Function.
3. Select the Psi Distribution function category.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>RNG</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binomial</td>
<td>PsiBinomial(n,p)</td>
<td>Returns the number of “successes” in a sample of size n where each trial has a probability p of “success.”</td>
</tr>
<tr>
<td>Discrete</td>
<td>PsiDiscrete({x_1,x_2,...,x_n},{p_1,p_2,...,p_n})</td>
<td>Returns one of the n values represented by the x_1. The value x_i occurs with probability p_i.</td>
</tr>
<tr>
<td>Discrete</td>
<td>PsiDisUniform ({x_1,x_2,...,x_n})</td>
<td>Returns one of the n values represented by the x_i. Each value x_i is equally likely to occur.</td>
</tr>
<tr>
<td>Poisson</td>
<td>PsiPoisson(λ)</td>
<td>Returns a random number of events occurring per some unit of measure (for example, arrivals per hour, defects per yard, and so on). The parameter λ represents the average number of events occurring per unit of measure.</td>
</tr>
<tr>
<td>Chi-square</td>
<td>PsiChiSquare(λ)</td>
<td>Returns a value from a chi-square distribution with mean λ.</td>
</tr>
<tr>
<td>Continuous</td>
<td>PsiUniform(min, max)</td>
<td>Returns a value in the range from a minimum (min) to a maximum (max). Each value in this range is equally likely to occur.</td>
</tr>
<tr>
<td>Exponential</td>
<td>PsiExponential(λ)</td>
<td>Returns a value from an exponential distribution with mean λ. Often used to model the time between events or the lifetime of a device with a constant probability of failure.</td>
</tr>
<tr>
<td>Normal</td>
<td>PsiNormal(μ,σ)</td>
<td>Returns a value from a normal distribution with mean μ and standard deviation σ.</td>
</tr>
<tr>
<td>Truncated Normal</td>
<td>PsiNormal(μ,σ,PsiTruncate(min,max))</td>
<td>Same as PsiNormal except the distribution is truncated to the range specified by a minimum (min) and a maximum (max).</td>
</tr>
<tr>
<td>Triangular</td>
<td>PsiTriangular(min, most likely, max)</td>
<td>Returns a value from a triangular distribution covering the range specified by a minimum (min) and a maximum (max). The shape of the distribution is then determined by the size of the most likely value relative to min and max.</td>
</tr>
</tbody>
</table>

*Commonly used RNGs supplied with Risk Solver Platform*
Chapter 12 Introduction to Simulation Using Risk Solver Platform

Examples of distributions associated with selected discrete RNGs.

**Figure 12.4**

Examples of distributions associated with selected continuous RNGs.

**Figure 12.5**
12.6.1 Discrete vs. Continuous Random Variables

An important distinction exists between the graphs in Figures 12.4 and 12.5. In particular, the RNGs depicted in Figure 12.4 generate discrete outcomes, whereas those represented in Figure 12.5 generate continuous outcomes. That is, some of the RNGs listed in Figure 12.3 can return only a distinct set of individual values, whereas the other RNGs can return any value from an infinite set of values. The distinction between discrete and continuous random variables is very important.

For example, the number of defective tires on a new car is a discrete random variable because it can assume only one of five distinct values: 0, 1, 2, 3, or 4. On the other hand, the amount of fuel in a new car is a continuous random variable because it can assume any value between 0 and the maximum capacity of the fuel tank. Thus, when selecting an RNG for an uncertain variable in a model, it is important to consider whether the variable can assume discrete or continuous values.

12.7 PREPARING THE MODEL FOR SIMULATION

To apply simulation to the model for Hungry Dawg Restaurants described earlier, we must first select appropriate RNGs for the uncertain variables in the model. If available, historical data on the uncertain variables could be analyzed to determine appropriate RNGs for these variables. Alternatively, the historical data itself can be sampled from using Risk Solver Platform’s PsiDisUniform, PsiResample, PsiSip or PsiSlurp functions. (Please refer to the Risk Solver Platform user manual for more information about these topics.) Risk Solver Platform also has the ability to automatically identify probability distributions that fit your historical data reasonably well. However, if past data are not available, or if we have some reason to expect the future behavior of a variable to be significantly different from the past, then we must use judgment in selecting appropriate RNGs to model the random behavior of the uncertain variables.

For our example problem, let’s assume that by analyzing historical data, we determined that the change in the number of covered employees from one month to the next is expected to vary uniformly between a 3% decrease and a 7% increase. (Note that this should cause the average change in the number of employees to be a 2% increase, because 0.02 is the midpoint between –0.03 and +0.07.) Further, assume that we can model the average monthly claim per covered employee as a normally distributed random variable with the mean (µ) increasing by 1% per month and a standard deviation (σ) of approximately $3. (Note that this will cause the average increase in claims per covered employee from one month to the next to be approximately 1%.) These assumptions are reflected in cells F5 through H6 at the top of Figure 12.6 (and in the file Fig12-6.xls).

To implement the formula to generate a random number of employees covered by the health plan, we will use the PsiUniform( ) function described in Figure 12.3. Because the change in the number of employees from one month to the next can vary between a 3% decrease and a 7% increase, the number of employees in the current month is equal to the number of employees in the previous month multiplied by the sum of 1 plus the percentage change. Applying this logic, we obtain the following equation for the number of employees in a given month:

\[
\text{Number of employees in current month} = \text{Number of employees in previous month} \times \text{PsiUniform}(0.97,1.07)
\]

If the PsiUniform( ) function returns the value 0.97, this formula causes the number of employees in the current month to equal 97% of the number in the previous month (a 3% decrease). Alternatively, if the PsiUniform( ) function returns the value 1.07, this formula causes the number of employees in the current month to equal 107% of the number in the previous month (a 7% increase). All the values between these two extremes (between 0.97 and 1.07) are also possible and equally likely to occur. The following formulas were used to create formulas that randomly generate the number of employees in each month in Figure 12.6:
Formula for cell B11: \[ =D5*\text{PsiUniform}(1-F5,1+H5) \]

Formula for cell B12: \[ =B11*\text{PsiUniform}(1-$F$5,1+$H$5) \]
(Copy to B13 through B22.)

Note that the terms “1–$F$5” and “1+$H$5” in the above formulas generate the values 0.97 and 1.07, respectively.

To implement the formula to generate the average claims per covered employee in each month, we will use the PsiNormal( ) function described in Figure 12.3. This formula requires that we supply the value of the mean (\( \mu \)) and standard deviation (\( \sigma \)) of the distribution from which we want to sample. The assumed $3 standard deviation (\( \sigma \)) for the average monthly claim, shown in cell H6 in Figure 12.6, is constant from month to month. Thus, the only remaining problem is to figure out the proper mean value (\( \mu \)) for each month.
In this case, the mean for any given month should be 1% larger than the mean in the previous month. For example, the mean for month 1 is:

\[ \text{Mean in month 1} = (\text{original mean}) \times 1.01 \]

and the mean for month 2 is:

\[ \text{Mean in month 2} = (\text{mean in month 1}) \times 1.01 \]

If we substitute the previous definition of the mean in month 1 into the previous equation, we obtain:

\[ \text{Mean in month 2} = (\text{original mean}) \times (1.01)^2 \]

Similarly, the mean in month 3 is:

\[ \text{Mean in month 3} = (\text{mean in month 2}) \times 1.01 = (\text{original mean}) \times (1.01)^3 \]

So in general, the mean (\( \mu \)) for month \( n \) is:

\[ \text{Mean in month } n = (\text{original mean}) \times (1.01)^n \]

Thus, to generate the average claim per covered employee in each month, we use the following formula:

Formula for cell D11:  \( =\text{PsiNormal}(\$D$6*(1+$F$6)^A11,\$H$6) \)

(Copy to D12 through D22.)

The term “\( \$D$6*(1+$F$6)^A11 \)” in this formula implements the general definition of the mean (\( \mu \)) in month \( n \).

After entering the appropriate RNGs, each time we press the recalculate key (the function key [F9]), the RNGs automatically select new values for all the cells in the spreadsheet that represent uncertain (or random) variables. Similarly, with each recalculation, a new value for the bottom-line performance measure (total company cost) appears in cell G23. Thus, by pressing the recalculate key several times, we can observe representative values of the company’s total cost for health claims. This also helps to verify that we implemented the RNGs correctly and that they are generating appropriate values for each uncertain cell.

12.7.1 Alternate RNG Entry

Risk Solver Platform also offers an alternate way of entering RNGs in spreadsheet models. If you select a worksheet cell and then click any of the icons in the Distributions icon on the Risk Solver Platform ribbon shown in Figure 12.3, a Distribution dialog appears similar to the one shown in Figure 12.7 (for the normal probability distribution with the dialog’s panels expanded). This dialog allows you to select any of the available probability distributions and see its shape as you vary the value of the various parameters. It also shows you the Psi function required to implement the RNG for the displayed probability distribution. Risk Solver Platform then automatically implements in your model the appropriate formula for the RNG you select. While this is a very useful feature of Risk Solver Platform, the RNG formula created by Risk Solver Platform often requires some manual editing to make it work correctly with the rest of the model you are building; particularly if you intend to copy this RNG formula to other cells in your workbook. (NOTE: Double clicking any cell containing a Psi distribution function also launches the type of dialog shown in Figure 12.7.)

Note that the Fit Distribution icon (resembling a bar chart, immediately to the left of the Expand Panels icon) at the top of Figure 12.7 launches the Fit Options dialog in Figure 12.8. If you have
historical data for one of the random variables in your model, you can use this dialog to instruct Risk Solver Platform to automatically fit a probability distribution to your data.

**Figure 12.7**

**Figure 12.8**
**12.8 RUNNING THE SIMULATION**

The next step in performing the simulation involves recalculating the spreadsheet several hundred or several thousand times and recording the resulting values generated for the output cell(s), or bottom-line performance measure(s). Fortunately, Risk Solver Platform can do this for us very easily if we indicate how many times we want it to replicate the model (or how many trials we want it to perform) and which cell(s) in the spreadsheet we want to track.

**12.8.1 Selecting the Output Cells to Track**

We can use the Add Output button on the Risk Solver Platform menu (see Figure 12.9) to indicate the output cell (or cells) that we want Risk Solver Platform to track during the simulation. In the current example, cell G23 represents the output cell we want Risk Solver Platform to track. To indicate this:

1. Click cell G23.
2. Click the Results icon on the Risk Solver Platform menu.
3. Click the “Output” option.
4. Click the “In Cell” option.

If you now look at the formula in cell G23, as shown in Figure 12.9, you will observe that it has been changed to:

Formula for cell G23: $=\text{SUM(G11:G22)} + \text{PsiOutput()}$

![Selecting the output cell to track.](Figure 12.9)
Clicking Risk Solver Platform’s Output, In Cell command with cell G23 selected caused the PsiOutput( ) function to be added to the original formula in cell G23. This is how Risk Solver Platform identifies output cells for our model. If you prefer, you can also manually add the PsiOutput( ) function to the contents of any numeric cell in your workbook to designate it as an output cell to Risk Solver. Alternatively, in any empty cell in the worksheet, we could enter the formula =PsiOutput(G23) and Risk Solver Platform would then know that cell G23 was an output cell. (This is also what happens if you choose the “Referred Cell” option rather than the “In Cell” option when using the Results, Output command.)

12.8.2 Selecting the Number of Replications

If we click the Options icon on the Risk Solver Platform menu shown in Figure 12.9, the Risk Solver Platform Options dialog box shown in Figure 12.10 appears. This dialog allows you to control several aspects of the simulation analysis. The Trials per Simulation option allows you to specify the number of trials or replications of your model Risk Solver Platform will generate when performing a simulation. All the examples in this book will use 1000 trials per simulation.

You might well wonder why we selected 1000 trials. Why not 500, or 10,000? Unfortunately, there is no easy answer to this question. Remember that the goal in simulation is to estimate various characteristics about the bottom-line performance measure(s) under consideration. For example, we might want to estimate the mean value of the performance measure and the shape of its probability distribution.
However, a different value of the bottom-line performance measure occurs each time we manually recalculate the model in Figure 12.9. Thus, there is an infinite number of possibilities—or an **infinite population**—of total company cost values associated with this model.

We cannot analyze all of these infinite possibilities. But by taking a large enough sample from this infinite population, we can make reasonably accurate estimates about the characteristics of the underlying infinite population of values. The larger the sample we take (that is, the more replications we do), the more accurate our final results will be. Although Risk Solver Platform is extremely fast, performing many replications takes time (especially for large models), so we must make a trade-off in terms of estimation accuracy versus convenience. Thus, there is no simple answer to the question of how many replications to perform, but, as a bare minimum you should always perform at least 500 replications, and more as time permits or accuracy demands.

**Software Note**
The educational version of Risk Solver Platform allows for only 1000 trials per simulation. The commercial version of this product removes this restriction and allows for as many trials as you desire.

### 12.8.3 Selecting What Gets Displayed on the Worksheet

When Risk Solver Platform carries out our simulation it generates 1000 replications or trials of our model. So for each Psi distribution and Psi output cell, Risk Solver Platform will compute and store 1000 values; but it can only display one value in any particular cell. So which of the 1000 values do we want it to display? Or might we prefer Risk Solver Platform to display the mean (average) of the 1000 values? Our answers to these questions can be communicated to Risk Solver Platform via the “Value to Display” setting on the Risk Solver Platform Options dialog in Figure 12.10 (or via the trial display counter immediately below the “Thaw” icon in the Tools section of the Risk Solver Platform ribbon shown in Figure 12.9). Using this option we can instruct Risk Solver Platform to display the value of one particular trial (the default setting) or have it display the mean of the sample of our replications.

It is important to note that if you select the sample mean option, Risk Solver Platform returns the sample mean for each Psi distribution cell and the resulting **computed** values for any Psi output cells. These computed values **may** or **may not** be the mean value of the Psi output cells depending on the nature of the functional relationship between the Psi distribution cells and the Psi output cells.

It is also important to note that if you ask Risk Solver Platform to display the values associated with one particular trial, the numbers displayed on the worksheet represent one random replication of your model that is no more special or important than any of the other replications in the simulation. Of course, any one random trial might be very unrepresentative of the typical values for the cells in the worksheet. As mentioned earlier, what we are really interested in is the **distribution** of outcomes associated with our output cells. As we will see, Risk Solver Platform offers a very elegant yet simple way to look at and answer questions about the distribution of outcomes associated with the output cells in a spreadsheet model.

### 12.8.4 Running The Simulation

Having identified the output cells to track and the number of replications to perform, we now need to instruct Risk Solver Platform to perform the simulation. This can be done in two different ways. The Simulate dropdown on the Risk Solver Platform ribbon has options for “Interactive” and “Run Once”. The “Run Once” option does just that – it runs one simulation consisting of the currently specified number of trials. Alternatively, you can select the “Interactive” option -- or simply click the Simulate icon (that looks like a light bulb) in the Risk Solver Platform menu. When the Simulate icon is on (or the light bulb is illuminated) Risk Solver Platform is in **interactive simulation** mode. In this mode, anytime you make a change to your workbook that requires the spreadsheet to be recalculated (or manually recalculate the spreadsheet by pressing the [F9] key), Risk Solver Platform performs a complete
simulation of your model—generating however many trials you specified per simulation in the Risk Solver Platform Options dialog in Figure 12.10. So while in interactive simulation mode, manually recalculating your workbook may actually cause your model to be replicated 1000 times. If this sounds like a lot of computational work, it is. However, if you use Risk Solver Platform’s internal spreadsheet interpreter (by choosing the “Use PSI Interpreter” option in the Risk Solver Platform Options dialog in Figure 12.10) these trials will usually be executed very quickly. The first time Risk Solver Platform performs a simulation for a given workbook it must parse or interpret the formulas in the spreadsheet, which sometimes takes a few seconds. However, after this is accomplished once, Risk Solver Platform performs future simulations with impressive speed.

12.9 DATA ANALYSIS

As mentioned earlier, the objective of performing a simulation is to estimate various characteristics of the outputs or bottom-line performance measures that are influenced by uncertainty in some or all of the input variables. While in interactive simulation mode, simply double click on any of the output cells in your model (identified using the PsiOutput( ) as described earlier) and Risk Solver Platform opens a dialog that allows you to summarize the output data in a variety of ways. Figure 12.11 shows the Risk Solver Platform detailed statistics window, created by double clicking cell G23 (representing the total company cost) in Figure 12.9.

![Summary statistics for the simulation trials.](Figure 12.11)
12.9.1 The Best Case and the Worst Case

As shown in Figure 12.11, the average (or mean) value for cell G23 is approximately $36.1 million. (If you are working through this example on a computer, the results you generate may be somewhat different from the results shown here because you may be working with a different sample of 1000 observations.) However, decision makers usually want to know the best-case and worst-case scenarios to get an idea of the range of possible outcomes they might face. This information is available from the simulation results, as shown by the Minimum and Maximum values listed in Figure 12.11.

Although the average total cost value observed in the 1000 replications is $36.1 million, in one case the total cost is approximately $29.3 million (representing the minimum or best case) and in another case the total cost is approximately $45.9 million (representing the maximum or worst case). These figures should give the decision maker a good idea about the range of possible cost values that might occur. Note that these values might be difficult to determine manually in a complex model with many uncertain independent variables.

12.9.2 Viewing the Distribution of the Output Cells

The best- and worst-case scenarios are the most extreme outcomes, and might not be likely to occur. To determine the likelihood of these outcomes requires that we know something about the shape of the distribution of our bottom-line performance measure. To view a graph of the distribution of the results of the 1000 replications generated for the output cell in our model, we can switch to the Frequency tab on the simulation results dialog for G23 as shown in Figure 12.12.

![Frequency distribution of the sampled total company costs.](image)

Figure 12.12 provides a visual summary of the approximate shape of the probability distribution associated with the output cell tracked by Risk Solver Platform during the simulation. In this case the shape of the distribution associated with the total cost variable is somewhat bell-shaped, with a maximum value around $46 million and a minimum value around $29 million. Thus, we now have a clear idea of the shape of the distribution associated with our bottom-line performance measure—one of the goals in simulation. You can right click on the frequency graphs like the one shown in Figure 12.12 to add upper
and lower markers to the graph and then click and drag these markers to identify the probability associated with different ranges of outcomes. In Figure 12.12 we see that approximately 95% of our trials resulted in outcomes with total cost values between $32 million and $40.6 million.

In this example, cell G23 (representing total company cost) is the only output cell that we identified (via the PsiOutput( ) function as discussed in section 12.8.1). However, it is important to note that if we tracked more than one output cell during the simulation (using multiple PsiOutput( ) functions), we could display histograms of the values occurring in these other output cells in a similar manner.

**Software Note**

You can right click on the frequency graphs like the one shown in Figure 12.12 to add upper and lower markers to the graph. You may then click and drag these markers to identify the probability associated with different ranges of outcomes.

### 12.9.3 Viewing the Cumulative Distribution of the Output Cells

At times, we might want to view a graph of the cumulative probability distribution associated with one of the output cells tracked during a simulation. For example, suppose that the chief financial officer (CFO) for Hungry Dawg would rather accrue an excess of money to pay health claims than not accrue enough money. The CFO might want to know what amount the company should accrue so that there is only a 10% chance of coming up short of funds at the end of the year. So, how much money would you recommend be accrued?

Figure 12.13 shows a graph of the cumulative probability distribution of the values that occurred in cell G23 during the simulation. This graph could help us ascertain and explain the answer to the preceding question.

![Cumulative frequency distribution of sampled total company costs.](image)

**Figure 12.13**
This graph displays the probability of the selected output cell taking on a value smaller than each value on the X-axis. For example, this graph indicates that approximately a 20% chance exists of the output cell (Total Company Cost) assuming a value smaller than approximately $34 million. Similarly, this graph indicates that roughly a 80% chance exists of total costs being less than approximately $38 million (or a 20% chance of total costs exceeding approximately $38 million). Thus, from this graph, we would estimate that roughly a 10% chance exists of the company’s costs exceeding approximately $39 million.

12.9.4 Obtaining Other Cumulative Probabilities

We can also answer the CFO’s question from information in the Percentiles tab shown in Figure 12.14. This window reveals a number of percentile values for the output cell G23. For example, the 75th percentile of the values generated for the output cell is $37,684,486 – or 75% of the 1000 values generated for cell G23 are less than or equal to this value. Similarly, the 90th percentile of the distribution of values is $39,038,823. Thus, based on these results, if the company accrues $39 million, we would expect that only about a 10% chance exists of the actual company costs exceeding this amount.
The ability to perform this type of analysis highlights the power and value of simulation and Risk Solver Platform. For example, how could we have answered the CFO’s question about how much money to accrue using best-case/worst-case analysis or what-if analysis? The fact is we could not answer the question with any degree of accuracy without using simulation.

12.9.5 Sensitivity Analysis

At times, you may be interested in examining how sensitive the simulation output results are to various uncertain cells in the model. This helps us determine which input (or uncertain) cells are most influential in effecting the bottom line output performance measure in the model. Such information can be useful as it can help direct our efforts to ensure that the most influential input cells are modeled accurately. In some cases, it can also help managers control (or reduce the variability in) the output variable by taking steps to reduce the variability in the most influential input variables.

Figure 12.15 shows how a Sensitivity chart identifies and summarizes the uncertain (RNG) cells in our model that are most significantly correlated (linearly) with the total company cost (output) values generated for cell G23. As this graph shows, the number of covered employees each month (in column B in our spreadsheet model) tends to have the largest impact on the total company cost.
12.10 THE UNCERTAINTY OF SAMPLING

To this point, we have used simulation to generate 1000 observations on our bottom-line performance measure and then calculated various statistics to describe the characteristics and behavior of the performance measure. For example, Figure 12.11 indicates that the mean company cost value in our sample is $36,132,938, and Figure 12.14 shows that a 90% chance exists of this performance measure assuming a value less than $39,038,823. But what if we repeat this process and generate another 1000 observations? Would the sample mean for the new 1000 observations also be exactly $36,132,938? Or would exactly 90% of the observations in the new sample be less than $39,038,823?

The answer to both these questions is “probably not.” The sample of 1000 observations used in our analysis was taken from a population of values that is theoretically infinite in size. That is, if we had enough time and our computer had enough memory, we could generate an infinite number of values for our bottom-line performance measure. Theoretically, we could then analyze this infinite population of values to determine its true mean value, its true standard deviation, and the true probability of the performance measure being less than $39,038,823. Unfortunately, we do not have the time or computer resources to determine these true characteristics (or parameters) of the population. The best we can do is take a sample from this population and, based on our sample, make estimates about the true characteristics of the underlying population. Our estimates will differ depending on the sample we choose and the size of the sample.

So, the mean of the sample we take is probably not equal to the true mean we would observe if we could analyze the entire population of values for our performance measure. The sample mean we calculate is just an estimate of the true population mean. In our example problem, we estimated that a 90% chance exists for our output variable to assume a value less than $39,038,823. However, this most likely is not equal to the true probability we would calculate if we could analyze the entire population. Thus, there is some element of uncertainty surrounding the statistical estimates resulting from simulation because we are using a sample to make inferences about the population. Fortunately, there are ways of measuring and describing the amount of uncertainty present in some of the estimates we make about the population under study. This is typically done by constructing confidence intervals for the population parameters being estimated.

12.11.1 Constructing a Confidence Interval for the True Population Mean

Constructing a confidence interval for the true population mean is a simple process. If \( \bar{y} \) and \( s \) represent, respectively, the mean and standard deviation of a sample of size \( n \) from any population, then assuming \( n \) is sufficiently large (\( n \geq 30 \)), the Central Limit Theorem tells us that the lower and upper limits of a 95% confidence interval for the true mean of the population are represented by:

\[
\text{95% Lower Confidence Limit} = \bar{y} - 1.96 \times \frac{s}{\sqrt{n}}
\]

\[
\text{95% Upper Confidence Limit} = \bar{y} + 1.96 \times \frac{s}{\sqrt{n}}
\]

Although we can be fairly certain that the sample mean we calculate from our sample data is not equal to the true population mean, we can be 95% confident that the true mean of the population falls between the lower and upper limits given previously. If we want a 90% or 99% confidence interval, we must change the value 1.96 in the previous equation to 1.645 or 2.575, respectively. The values 1.645, 1.96, and 2.575 represent the 95, 97.5, and 99.5 percentiles of the standard normal distribution. Any percentile of the standard normal distribution can be obtained using Excel's NORMSINV( ) function.

For our example, the lower and upper limits of a 95% confidence interval for the true mean of the population of total company cost values can be calculated easily, as shown in cells B9 and B10 in Figure 12.16.

Formula for cell B9: \( =B4-NORMSINV(1-B7/2)*B5/SQRT(B6) \)

Formula for cell B10: \( =B4+NORMSINV(1-B7/2)*B5/SQRT(B6) \)
Thus, we can be 95% confident that the true mean of the population of total company cost values falls somewhere in the interval from $35,992,789 to $36,273,087.

Notice that the sample mean and standard deviation shown in cells B4 and B5 of Figure 12.16 can be obtained directly from the simulation results using two of Risk Solver Platform’s Psi statistics functions.

Formula for cell B4:  \( =\text{PsiMean}('\text{Health Claims Model}'!G23) \)

Formula for cell B5:  \( =\text{PsiStdDev}('\text{Health Claims Model}'!G23) \)

These formulas, respectively, return the mean and standard deviation of the 1000 numbers that Risk Solver Platform has stored for cell G23. The Statistic, Measure, and Range icons on the Risk Solve menu provide galleries of several other Psi functions that can be used in a similar way to calculate and report simulation results directly on a spreadsheet. These functions can be extremely helpful in summarizing simulation results. However, it is also important to note that these functions can only work while Risk Solver Platform is in interactive simulation mode.

### 12.11.2 Constructing a Confidence Interval for a Population Proportion

In our example, we estimated that 90% of the population of total company cost values fall below $39,038,823 based on our sample of 1000 observations. However, if we could evaluate the entire population of total cost values, we might find that only 80% of these values fall below $39,038,823. Or, we might find that 99% of the entire population falls below this mark. It would be helpful to determine how accurate the 90% value is. So, at times, we might want to construct a confidence interval for the true proportion of a population that falls below (or above) some value, for example \( Y_p \).
To see how this is done, let \( p \) denote the proportion of observations in a sample of size \( n \) that falls below some value \( Y_p \). Assuming that \( n \) is sufficiently large (\( n \geq 30 \)), the Central Limit Theorem tells us that the lower and upper limits of a 95% confidence interval for the true proportion of the population falling below \( Y_p \) are represented by:

\[
\text{95\% Lower Confidence Limit} = p - 1.96 \times \sqrt{\frac{p(1-p)}{n}} \\
\text{95\% Upper Confidence Limit} = p + 1.96 \times \sqrt{\frac{p(1-p)}{n}}
\]

Although we can be fairly certain that the proportion of observations falling below \( Y_p \) in our sample is not equal to the true proportion of the population falling below \( Y_p \), we can be 95% confident that the true proportion of the population falling below \( Y_p \) is contained within the lower and upper limits given previously. Again, if we want a 90% or 99% confidence interval, we must change the value 1.96 in the previous equations to 1.645 or 2.575, respectively.

Using these formulas, we can calculate the lower and upper limits of a 95% confidence interval for the true proportion of the population falling below \$39,038,823.\ From our simulation results we know that 90% of the observations in our sample are less than \$39,038,823. Thus, our estimated value of \( p \) is 0.90. This value was entered into cell B12 in Figure 12.16. The upper and lower limits of a 95% confidence interval for the true proportion of the population falling below \$39,038,823 are calculated in cells B13 and B14 of Figure 12.16 using the following formulas:

\[
\text{Formula for cell B13: } =E12 - \text{NORMSINV}(1 - B7/2) \times \text{SQRT}(B12 \times (1 - B12) / B6) \\
\text{Formula for cell B14: } = E12 + \text{NORMSINV}(1 - B7/2) \times \text{SQRT}(B12 \times (1 - B12) / B6)
\]

We can be 95% confident that the true proportion of the population of total cost values falling below \$39,038,823 is between 0.881 and 0.919. Because this interval is fairly tight around the value 0.90, we can be reasonably certain that the \$39,038,823 figure quoted to the CFO has approximately a 10% chance of being exceeded.

### 12.11.3 Sample Sizes and Confidence Interval Widths

The formulas for the confidence intervals in the previous section depend on the number of replications (\( n \)) in the simulation. As the number of replications (\( n \)) increases, the width of the confidence interval decreases (or becomes more precise). Thus, for a given level of confidence (for example, 95%), the only way to make the upper and lower limits of the interval closer together (or tighter) is to make \( n \) larger—that is, use a larger sample size. A larger sample should provide more information about the population and, therefore, allow us to be more accurate in estimating the true parameters of the population.

### 12.12 THE BENEFITS OF SIMULATION

So what have we accomplished through simulation? Are we really better off than if we had just used the results of the original model proposed in Figure 12.2? The estimated value for the expected total cost to the company in Figure 12.2 is comparable to that obtained through simulation (although this might not always be the case). But remember that the goal of modeling is to give us greater insight into a problem to help us make more informed decisions.

The results of our simulation do give us greater insight into the problem. In particular, we now have some idea of the best- and worst-case total cost outcomes for the company. We have a better idea of the distribution and variability of the possible outcomes, and a more precise idea about where the mean of the distribution is located. We also now have a way of determining how likely it is for the actual outcome to fall above or below some value. Thus, in addition to our greater insight and understanding of the problem, we also have solid empirical evidence (the facts and figures) to support our recommendations.
Applying Simulation in Personal Financial Planning

A recent article in the *Wall Street Journal* highlighted the importance of simulation in evaluating the risk in personal financial investments. "Many people are taking a lot more risk than they realize. They are walking around with a false sense of security," said Christopher Cordaro, an investment adviser at Bugen Stuart Korn & Cordaro of Chatman, N.J. Using one on-line retirement calculator, after entering information about your finances and assumptions about investment gains, the screen shows a green or red light indicating whether you have saved enough to retire. What is doesn't tell you is whether there is a 95% chance or just a 60% chance that your plan will succeed.

Although any planning is better than nothing, traditional planning models spit out answers that create "the illusion that the number is a certainty, when it isn't," said Ross Levin, a Minneapolis investment adviser who uses simulation. Mutual-fund firm T. Rowe Price applied simulation to its recently launched Retirement Income Manager, a personalized consultation service that helps retirees understand how much income they can afford without outliving their assets or depleting funds they want to leave to heirs. A number of independent financial planning companies are now also using simulation software.

In the face of widely divergent possibilities, the number crunching involved in simulation can bring some peace of mind. Steven Haas, of Randolph, N.J., was not sure whether he had saved enough to accept an early retirement package from AT&T at age 53. After his financial advisor used simulation to determine he had a 95% probability that his money would last until he reached age 110, Mr. Haas took the package and retired, reporting "I found that even under significantly negative scenarios we could live comfortably. It relieved a lot of anxiety."

Adapted from "Monte Carlo Financial Simulator May Be A Good Bet For Planning," *Wall Street Journal*, Section C1, April 27, 2000 by Karen Hube.

12.13 ADDITIONAL USES OF SIMULATION

Earlier, we indicated that simulation is a technique that describes the behavior or characteristics of a bottom-line performance measure. The next several examples show how describing the behavior of a performance measure gives a manager a useful tool in determining the optimal value for one or more controllable parameters in a decision problem. These examples reinforce the mechanics of using simulation and also demonstrate some additional capabilities of Risk Solver Platform.

12.14 A RESERVATION MANAGEMENT EXAMPLE

Businesses that allow customers to make reservations for services (such as airlines, hotels, and car rental companies) know that some percentage of the reservations made will not be used for one reason or another, leaving these companies with a difficult decision problem. If they accept reservations for only the number of customers that can actually be served, then a portion of the company’s assets will be underutilized when some customers with reservations fail to arrive. On the other hand, if they overbook (or accept more reservations than can be handled), then at times, more customers will arrive than can actually be served. This typically results in additional financial costs to the company and often generates ill-will among those customers who cannot be served. The following example illustrates how simulation might be used to help a company determine the optimal number of reservations to accept.

Marty Ford is an operations analyst for Piedmont Commuter Airlines (PCA). Recently, Marty was asked to make a recommendation on how many reservations PCA should book on Flight 343—a flight from a small regional airport in New England to a major hub at Boston’s Logan airport. The plane used on Flight 343 is a small twin-engine turbo-prop with 19 passenger seats available. PCA sells nonrefundable tickets for Flight 343 for $150 per seat.
Industry statistics show that for every ticket sold for a commuter flight, a 0.10 probability exists that the ticket holder will not be on the flight. Thus, if PCA sells 19 tickets for this flight, there is a fairly good chance that one or more seats on the plane will be empty. Of course, empty seats represent lost potential revenue to the company. On the other hand, if PCA overbooks this flight and more than 19 passengers show up, some of them will have to be bumped to a later flight.

To compensate for the inconvenience of being bumped, PCA gives these passengers vouchers for a free meal, a free flight at a later date, and sometimes also pays for them to stay overnight in a hotel near the airport. PCA pays an average of $325 (including the cost of lost goodwill) for each passenger that gets bumped. Marty wants to determine if PCA can increase profits by overbooking this flight and, if so, how many reservations should be accepted to produce the maximum average profit. To assist in the analysis, Marty analyzed market research data for this flight that reveals the following probability distribution of demand for this flight:

<table>
<thead>
<tr>
<th>Seats Demanded</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>0.03</td>
</tr>
<tr>
<td>15</td>
<td>0.05</td>
</tr>
<tr>
<td>16</td>
<td>0.07</td>
</tr>
<tr>
<td>17</td>
<td>0.09</td>
</tr>
<tr>
<td>18</td>
<td>0.11</td>
</tr>
<tr>
<td>19</td>
<td>0.15</td>
</tr>
<tr>
<td>20</td>
<td>0.18</td>
</tr>
<tr>
<td>21</td>
<td>0.14</td>
</tr>
<tr>
<td>22</td>
<td>0.08</td>
</tr>
<tr>
<td>23</td>
<td>0.05</td>
</tr>
<tr>
<td>24</td>
<td>0.03</td>
</tr>
<tr>
<td>25</td>
<td>0.02</td>
</tr>
</tbody>
</table>

12.14.1 Implementing the Model

A spreadsheet model for this problem is shown in Figure 12.17 (and in the file Fig12-17.xls). The spreadsheet begins by listing the relevant data from the problem, including the number of seats available on the plane, the price PCA charges for each seat, the probability of a no-show (a ticketed passenger not arriving in time for the flight), the cost of bumping passengers, and the number of reservations that will be accepted.

The distribution of demand for seats on the flight is summarized in columns E and F. Using this data, the number of seats demanded for a particular flight is randomly generated in cell C10 as follows:

Formula for cell C10: =PsiDiscrete(E5:E16, F5:F16)

The number of tickets actually sold for a flight cannot exceed the number of reservations the company is willing to accept. Thus, the number of tickets sold is calculated in cell C11 as follows:

Formula for cell C11: =MIN(C10,C8)

Because each ticketed passenger has a 0.10 probability of being a no-show, a 0.9 probability exists that each ticketed passenger will arrive in time to board the flight. Thus, the PsiBinomial( ) function (described in Figure 12.3) is used in cell C12 to model the number of ticketed passengers that actually arrive for a flight:

Formula for cell C12: =PsiBinomial(C11,1–C6)

Cell C14 represents the ticket revenue PCA earns based on the number of tickets it sells for each flight. The formula for this cell is:

Formula for cell C14: =C11*C5

Cell C15 computes the costs PCA incurs when passengers must be bumped (i.e., when the number of passengers wanting to board exceeds the number of available seats).

Formula for cell C15:  = MAX(C12–C4,0)*C7
Finally, cell C16 computes the marginal profit PCA earns on each flight. This is also the output cell to be tracked when simulating this model.

Formula for cell C16: \[ =C14-C15+\text{PsiOutput( )} \]

---

**Spreadsheet model for the overbooking problem.**

**Figure 12.17**

### 12.14.2 Details for Multiple Simulations

Marty wants to determine the number of reservations to accept that, on average, will result in the highest marginal profit. To do so, he needs to use the `PsiSimParam( )` function to simulate what would happen if 19, 20, 21, 22, 23, 24, and 25 reservations are accepted. Cell C8 contains the following formula:

Formula for cell C8: \[ =\text{PsiSimParam}(19,20,21,22,23,24,25) \]

This formula, along with the “Simulations to Run” settings shown in the Risk Solver Platform Options dialog in Figure 12.18, instructs Risk Solver Platform to use seven different values in cell C8 and simulate what will happen with each value.

When comparing different values for one or more decision variables, it is best if each possible value is evaluated in a simulation using exactly the same series of random numbers. In this way, any difference in the performance of two possible solutions can be attributed to the decision variables’ values and not the
result of a more favorable set of random numbers for one of the simulations. The "Sim. Random Seed" option shown in Figure 12.18 controls this behavior in Risk Solver Platform. By default, Risk Solver Platform will use a randomly chosen seed value to initialize its random number generator when performing multiple simulations using the PsiSimParam( ) function. Alternatively, you may override Risk Solver Platform’s default behavior and instruct it to use a seed value you specify when it performs multiple simulations. Choosing your own seed allows you to repeat the same simulation again in the future if needed.

It is worth noting that the "Sampling Method" options shown in Figure 12.18 also have an impact on the accuracy of the results of a simulation run. Using the “Monte Carlo” option, Risk Solver Platform is free to select any value for a particular RNG during each replication of the model. For example, Risk Solver Platform might repeatedly generate several very extreme (and rare!) values from the upper tail of a normal distribution. The “Latin Hypercube” option guards against this by ensuring that a fair representation of values is generated from the entire distribution for each RNG. As you might imagine, the Latin Hypercube sampling option requires a bit more work during each replication of the model, but it tends to generate more accurate simulation results in a fewer number of trials. Refer to Risk Solver Platform’s user manual for additional information about its supported sampling methods.

12.14.3 Running the Simulations

In Figure 12.17 we included the PsiOutput( ) function in our formula for cell C16 (representing marginal profit) to indicate it is the output cell we want Risk Solver Platform to track. In Figure 12.18 we indicated Risk Solver Platform should perform seven simulations (one for each value from 19 to 25 indicated by the PsiSimParam( ) function in cell C8) and run a simulation consisting of 1000 replications for each possible value for C8. If Risk Solver Platform is in interactive simulation mode, it will automatically carry out each of the desired seven simulations (involving 7,000 replications of our model).
12.14.3 Data Analysis

Risk Solver Platform provides a number of ways for us to look at the results of the seven simulations. If you double click on cell C16 (which includes the PsiOutput( ) function) we can look at the marginal profit results for each of the seven simulations. Figure 12.19 shows how Risk Solver Platform allows us to view the statistics associated with any of the seven simulations.

Using the Frequency tab, we can also simultaneously chart the cumulative frequency distribution of the marginal profit values associated with each of the seven simulations as shown in Figure 12.20.
Of course, we can also use Psi statistic functions to create a custom summary of the simulation results directly in a worksheet. An example of this is given in Figure 12.21. In the case of multiple simulations, note that the last argument of each Psi statistic function indicates to which set of simulation data the function applies. For example, the formula =PsiMean(Model!C16,1) would return the mean marginal profit from simulation one (where 19 reservations were accepted) while =PsiMean(Model!C16,5) would return the mean marginal profit from simulation five (where 23 reservations were accepted). This data makes it clear that if PCA wants to maximize its expected (or average) marginal profit, it should accept 21 reservations per flight. Accepting more than 21 reservations makes it possible to achieve higher levels of profit on some flights but, on average (over a large number of flights), accepting more than 21 reservations would result in less profit for the company if the assumptions in our model are correct.

Summary of results from all seven simulations.

**Figure 12.21**

### 12.15 AN INVENTORY CONTROL EXAMPLE

According to the *Wall Street Journal*, U.S. businesses recently had a combined inventory worth $884.77 billion dollars. Because so much money is tied up in inventories, businesses face many important decisions regarding the management of these assets. Frequently asked questions regarding inventory include:
• What’s the best level of inventory for a business to maintain?
• When should goods be reordered (or manufactured)?
• How much safety stock should be held in inventory?

The study of inventory control principles is split into two distinct areas—one assumes that demand is known (or deterministic), and the other assumes that demand is random (or stochastic). If demand is known, various formulas can be derived that provide answers to the previous questions (an example of one such formula is given in the discussion of the EOQ model in chapter 8). However, when demand for a product is uncertain or random, answers to the previous questions cannot be expressed in terms of a simple formula. In these situations, the technique of simulation proves to be a useful tool, as illustrated in the following example.

Laura Tanner is the owner of Millennium Computer Corporation (MCC), a retail computer store in Austin, Texas. Competition in retail computer sales is fierce—both in terms of price and service. Laura is concerned about the number of stockouts occurring on a popular type of computer monitor. Stockouts are very costly to the business because when customers cannot buy this item at MCC, they simply buy it from a competing store and MCC loses the sale (there are no backorders). Laura measures the effects of stockouts on her business in terms of service level, or the percentage of total demand that can be satisfied from inventory.

Laura has been following the policy of ordering 50 monitors whenever her daily ending inventory position (defined as ending inventory on hand plus outstanding orders) falls below her reorder point of 28 units. Laura places the order at the beginning of the next day. Orders are delivered at the beginning of the day and, therefore, can be used to satisfy demand on that day. For example, if the ending inventory position on day 2 is less than 28, Laura places the order at the beginning of day 3. If the actual time between order and delivery, or lead time, turns out to be four days, then the order arrives at the start of day 7. The current level of on-hand inventory is 50 units and no orders are pending.

MCC sells an average of six monitors per day. However, the actual number sold on any given day can vary. By reviewing her sales records for the past several months, Laura determined that the actual daily demand for this monitor is a random variable that can be described by the following probability distribution:

<table>
<thead>
<tr>
<th>Units Demanded</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.01</td>
<td>0.02</td>
<td>0.04</td>
<td>0.06</td>
<td>0.09</td>
<td>0.14</td>
<td>0.18</td>
<td>0.22</td>
<td>0.16</td>
<td>0.06</td>
<td>0.02</td>
</tr>
</tbody>
</table>

The manufacturer of this computer monitor is located in California. Although it takes an average of four days for MCC to receive an order from this company, Laura has determined that the lead time of a shipment of monitors is also a random variable that can be described by the following probability distribution:

<table>
<thead>
<tr>
<th>Lead Time (days)</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.2</td>
<td>0.6</td>
<td>0.2</td>
</tr>
</tbody>
</table>

One way to guard against stockouts and improve the service level is to increase the reorder point for the item so that more inventory is on hand to meet the demand occurring during the lead time. However, there are holding costs associated with keeping more inventory on hand. Laura wants to evaluate her current ordering policy for this item and determine if it might be possible to improve the service level without increasing the average amount of inventory on hand.

12.15.1 Creating the RNGs

To solve this problem, we need to build a model to represent the inventory of computer monitors during an average month of 30 days. This model must account for the random daily demands that can occur and
the random lead times encountered when orders are placed. First, we will consider how to create RNGs to model the daily demands and order lead times. The data for these variables are entered in the spreadsheet as shown in Figure 12.22 (and in the file Fig12-22.xls).

![RNG data for MCC's inventory problem.](Figure 12.22)

The order lead time and daily demand variables are both examples of *general, discrete* random variables because the possible outcomes they assume consist solely of integers, and the probabilities associated with each outcome are not equal (or not uniform). Thus, using the RiskDiscrete( ) function described in Figure 12.5, the RNGs for each variable are:

- RNG for order lead time: \( \text{PsiDiscrete(Data!$C$7:$C$9,Data!$D$7:$D$9)} \)
- RNG for daily demand: \( \text{PsiDiscrete(Data!$F$7:$F$17,Data!$G$7:$G$17)} \)

### 12.15.2 Implementing the Model

Now that we have a way of generating the random numbers needed in this problem, we can consider how the model should be built. Figure 12.22 shows the model representing 30 days of inventory activity. Notice that cells M5 and M6 have been reserved to represent, respectively, the reorder point and order quantity for the model.

The inventory on hand at the beginning of each day is calculated in column B in Figure 12.22. The beginning inventory for each day is simply the ending inventory from the previous day. The formulas in column B are:

- Formula for cell B6: \( =50 \)
- Formula for cell B7: \( =F6 \)

(Copy to B8 through B35.)

Column C represents the number of units scheduled to be received each day. We will discuss the formulas in column C after we discuss columns H, I, and J, which relate to ordering and order lead times.
Chapter 12 Introduction to Simulation Using Risk Solver Platform

Key Cell Formulas

<table>
<thead>
<tr>
<th>Cell</th>
<th>Formula</th>
<th>Copied to</th>
</tr>
</thead>
<tbody>
<tr>
<td>B6</td>
<td>=50</td>
<td>--</td>
</tr>
<tr>
<td>B7</td>
<td>=F6</td>
<td>B8:B35</td>
</tr>
<tr>
<td>C7</td>
<td>=COUNTIF($J$6:J6,A7)*$M$6</td>
<td>C8:C35</td>
</tr>
<tr>
<td>D6</td>
<td>=PsiDiscrete(Data!$F$7:$F$17,Data!$G$7:$G$17)</td>
<td>D7:D35</td>
</tr>
<tr>
<td>E6</td>
<td>=MIN(D6,B6+C6)</td>
<td>E7:E35</td>
</tr>
<tr>
<td>F6</td>
<td>=B6+C6-E6</td>
<td>F7:F35</td>
</tr>
<tr>
<td>G6</td>
<td>=F6</td>
<td>--</td>
</tr>
<tr>
<td>G7</td>
<td>=G6-E7+IF(H6=1,$M$6,0)</td>
<td>G8:G35</td>
</tr>
<tr>
<td>H6</td>
<td>=IF(G6&lt;$M$6,1,0)</td>
<td>H7:H35</td>
</tr>
<tr>
<td>I6</td>
<td>=IF(H6=0,0,PsiDiscrete(Data!$C$7:$C$9,Data!$D$7:$D$9))</td>
<td>I7:I35</td>
</tr>
<tr>
<td>J6</td>
<td>=IF(I6=0,0,A6+1+I6)</td>
<td>J7:J35</td>
</tr>
<tr>
<td>M9</td>
<td>=SUM(E6:E35)/SUM(D6:D35) + PsiOutput( )</td>
<td>--</td>
</tr>
<tr>
<td>M10</td>
<td>=AVERAGE(B6:B35) + PsiOutput( )</td>
<td>--</td>
</tr>
</tbody>
</table>

Spreadsheet for MCC's inventory problem.

Figure 12.23

In column D, we use the technique described earlier to generate random daily demands, as:

Formula for cell D6: 

\[ =\text{PsiDiscrete(Data!}F7:F17,\text{Data!}G7:G17) \]

(Copy to D7 through D35.)
Because it is possible for demand to exceed the available supply, column E indicates how much of the daily demand can be met. If the beginning inventory (in column B) plus the ordered units received (in column C) is greater than or equal to the actual demand, then all the demand can be satisfied; otherwise, MCC can sell only as many units as are available. This condition is modeled as:

Formula for cell E6: \text{MIN(D6,B6+C6)}
(Copy to E7 through E35.)

The values in column F represent the on-hand inventory at the end of each day, and are calculated as:

Formula for cell F6: \text{B6+C6–E6}
(Copy to F7 through F35.)

To determine whether to place an order, we first must calculate the inventory position, which was defined earlier as the ending inventory plus any outstanding orders. This is implemented in column G as:

Formula for cell G6: \text{F6}
Formula for cell G7: \text{G6–E7+IF(H6=1,$M$6,0)}
(Copy to G8 through G35.)

Column H indicates if an order should be placed based on inventory position and the reorder point, as:

Formula for cell H6: \text{IF(G6<$M$5,1,0)}
(Copy to H7 through H35.)

If an order is placed, then we must generate the random lead time required to receive the order. This is done in column I as:

Formula for cell I6: \text{IF(H6=0,0,PsiDiscrete(Data!$C$7:$C$9,Data!$D$7:$D$9))}
(Copy to I7 through I35.)

This formula returns the value 0 if no order was placed (if H6=0); otherwise, it returns a random lead time value (if H6=1).

If an order is placed, column J indicates the day on which the order will be received based on its random lead time in column I. This is done as:

Formula for cell J6: \text{IF(I6=0,0,A6+1+I6)}
(Copy to J7 through J35.)

The values in column C are coordinated with those in column J. The nonzero values in column J indicate the days on which orders will be received. For example, cell J9 indicates that an order will be received on day 10. The actual receipt of this order is reflected by the value of 50 in cell C15, which represents the receipt of an order at the beginning of day 10. The formula in cell C15 that achieves this is:

Formula for cell C15: \text{COUNTIF($J$6:J14,A15)*$M$6}

This formula counts how many times the value in cell A15 (representing day 10) appears as a scheduled receipt day between days 1 through 9 in column J. This represents the number of orders scheduled to be received on day 10. We then multiply this by the order quantity (50), given in cell M6 to determine the total units to be received on day 10. Thus, the values in column C are generated as:

Formula for cell C6: \text{=0}
Formula for cell C7: \text{COUNTIF($J$6:J6,A7)*$M$6}
(Copy to C8 through C35.)
The service level for the model is calculated in cell M9 using the values in columns D and E as:

Formula for cell M9: \( \frac{\text{SUM}(E6:E35)}{\text{SUM}(D6:D35)} + \text{PsiOutput}() \)

Again, the service level represents the proportion of total demand that can be satisfied from inventory and is one of the output cells we want Risk Solver Platform to track as we simulate this inventory system. The value in cell M9 indicates that in the scenario shown, 86.84% of the total demand is satisfied.

The average inventory level is also an output we want Risk Solver Platform to track. It is calculated in cell M10 by averaging the values in column B. This is accomplished as follows:

Formula for cell M10: \( \text{AVERAGE}(B6:B35) + \text{PsiOutput}() \)

### 12.15.3 Replicating the Model

The model in Figure 12.23 indicates one possible scenario that could occur if Laura uses a reorder point of 28 units for the computer monitor. Figures 12.24 and 12.25 show the results of using Risk Solver Platform to replicate this model 1000 times, tracking the values in cells M9 (service level) and M10 (average inventory) as output cells.

Figures 12.24 and 12.25 indicate that MCC’s current reorder point (28 units) and order quantity (50 units) results in an average service level of approximately 96% (with a minimum value around 83% and a maximum value of 100%) and an average inventory level of almost 26 monitors (with a minimum value around 20 and a maximum value near 33).
12.15.4 Optimizing the Model

Now suppose Laura wants to determine a reorder point and order quantity that provides an average service level of 98% while keeping the average inventory level as low as possible. One way to do this is to run additional simulations at various reorder point and inventory level combinations trying to find the combination of settings that produce the desired behavior. However, as you might imagine, this could be very time-consuming. Fortunately, Risk Solver Platform can solve this type of problem for us.

Risk Solver Platform provides access to the same heuristic evolutionary optimization engine available in Solver (as described in chapter 8). Like Solver, Risk Solver Platform allows us to maximize or minimize a value associated with some target or objective cell in a worksheet by changing the values of other cells (representing controllable decision variables) while satisfying various constraints. However, because the worksheet contains RNGs in various cells, to evaluate the behavior or quality of a particular solution Risk Solver Platform must simulate (or run multiple replications of) the model at each solution it considers. While this is very computationally intensive, Risk Solver Platform’s interactive simulation abilities allow these computations to be done quite rapidly.

Another difference between Risk Solver Platform and Solver is that when attempting to optimize a simulation model (also known as simulation optimization) we typically want to maximize or minimize the average value of (or some other statistic describing) the cell representing the objective or bottom-line performance measure. Again, this is due to the fact that there is not one definite or certain outcome associated with a particular solution in a simulation model; rather, there is a distribution of possible outcomes. Similarly, constraints are typically expressed as some statistical measure (e.g., average, percentile, standard deviation) of the constraint cell in question. So, in simulation optimization the goal is to automatically identify a solution (values for the decision variables) that causes a model of a process containing randomness (or uncertainty) to behave in the most desirable way possible.

Inventory results of 1000 replications of the MCC model.
Figure 12.26 (and the file Fig12-26.xls on your data disk) shows how the spreadsheet was changed to find the optimal solution to MCC’s inventory problem. Recall that Laura wants to determine the reorder point and order quantity that will keep the average inventory level as low as possible while achieving an average service level of 98%. To do this, we added formulas in cells M13 and M14 that compute, respectively, the average service level and average inventory level for the entire simulation as follows,

Formula for cell M13: =PsiMean(M9)
Formula for cell M14: =PsiMean(M10)

Let’s take a moment to make sure you understand the difference between the values in cells M9 and M13 and also between M10 and M14. In Figure 12.26, cells M9 and M10 are displaying, respectively, the service level and average inventory for the single replication of the model that is displayed. However, because each of these cells serve as output cells for the simulation (via the PsiOutput( ) functions shown in their formula definitions in Figure 12.23), when Risk Solver Platform is in interactive simulation mode (with the “Simulate” light bulb illuminated) there are actually 1000 values saved for cells M9 and M10, corresponding to 1000 trials of a simulation. So, in interactive simulation mode, the PsiMean( ) functions in cells M13 and M14 computed, respectively, the averages of the 1000 trial values associated with cells M9 and M10. (If you exit interactive simulation mode by “turning off” the Simulate light bulb, the
PsiMean( ) functions in cells M13 and M14 return the value “#N/A”). The values in cells M13 and M14 are the ones of interest from an optimization perspective because Laura is interested in the average service level and average inventory level over an entire 1000 trial simulation – not the average service level and average inventory level for any one particular trial. (The point being made in this paragraph is key to understanding simulation optimization, so be sure you understand this before proceeding.)

Clicking the “Model” icon on the Risk Solver Platform ribbon causes the Solver Options and Model Specifications pane to appear, as shown in Figure 12.26. This pane provides an integrated approach to optimization and simulation. Note that the “Simulation” section of this pane summarizes everything that Risk Solver Platform understands about the model in this spreadsheet: that cells D6 through D35 and I6 through I35 are uncertain (or random) variables, cells M9 and M10 are uncertain functions (outputs), and M13 and M14 are statistic functions (computing descriptive statistics about the simulation).

The “Optimization” section of the Solver Options and Model Specifications pane allows us to do everything that Solver can do – and more. In this case, we want to instruct Risk Solver Platform to minimize the average inventory in the simulation (in cell M14) by changing the values of the reorder point and order quantity (decision variables) in cells M5 and M6, respectively, while simultaneously keeping the simulation’s average service level (in cell M13) at or above 98%.

To specify the objective for this problem,

1. Select cell M14 (representing the average inventory for the simulation).
2. Select the word “Objective” in the Solver Options and Model Specifications pane.
3. Click on the green plus (“+”) symbol (indicated by a red circle in Figure 12.26).

The results of the above steps are shown in Figure 12.27. Note that after the objective is added, the bottom of the Risk Solver Platform pane displays various options associated with our action and we can indicate our desire to minimize the objective.
Next, we specify the adjustable (or changing) cells representing the decisions about the reorder point and order quantity. To do this,

1. Select cells M5 and M6 (representing the reorder point and order quantity, respectively).
2. Select “Normal” in the “Variables” section of the Solver Options and Model Specifications pane.
3. Click on the green plus (“+”) symbol.

The results of the above steps are shown in Figure 12.28. Note that additional variables (when needed) would be added in a similar manner. And after variables are added, the bottom of the Risk Solver Platform pane displays various options associated with the selected variables.

![Figure 12.28](image)

Defining the decision variables for the optimization model.

Next, we need to specify any constraints that apply to the problem. One such constraint relates to Laura’s desire to achieve at least a 98% average service level. To do this,

1. Select cell M13 (representing the average service level for the simulation).
2. Select “Normal” in the “Constraints” section of the Solver Options and Model Specifications pane.
3. Click on the green plus (“+”) symbol.

This results in the dialog shown in Figure 12.29 where we can indicate that cell M13 must be greater than or equal to 98% (or 0.98).
Defining the service level constraint for the optimization model.

Figure 12.29

After clicking “OK” in the dialog shown in Figure 12.29, we also need to add upper and lower bounds on the decision variables for this problem. We will assume that Laura is interested in considering values between 10 and 70 for both the reorder point and order quantity variables (cells M5 and M6). To create this constraint,

1. Select cells M5 and M6 (representing the reorder point and order quantity, respectively).
2. Select “Bound” in the “Constraints” section of the Solver Options and Model Specifications pane.
3. Click on the green plus (“+”) symbol.

Figure 12.30 shows the resulting dialog and settings to specify a lower bound of 10 for the decision variables. The same step of steps can be used to define an upper bound of 70, as shown in Figure 12.31.
Defining an upper bound for the decision variables.

Figure 12.31

Finally, we need to indicate that the decision variables may only take on integer values. To do this,

1. Select cells M5 and M6 (representing the reorder point and order quantity, respectively).
2. Select “Integers” in the “Constraints” section of the Solver Options and Model Specifications pane.
3. Click on the green plus (“+”) symbol.

Figure 12.32 shows the resulting dialog where we select the “int” option from the dropdown list to indicate that cells M5 and M6 must be integers.

Defining integer conditions for the decision variables.

Figure 12.32
Figure 12.33 shows a summary of the Risk Solver Platform settings required for the MCC problem. Clicking the green triangle (play) icon in the Solver Options and Model Specifications pane causes Risk Solver Platform to solve the problem.

Summary of Risk Solver Platform settings.

Remember that a separate simulation must be run for each combination of the decision variables that Risk Solver Platform chooses. Risk Solver Platform uses a number of heuristics to search intelligently for the best combination of decision variables. However, this is still inherently a very computationally intensive and time-consuming process and very complicated models could take hours (or days) of solution time.

As shown in Figure 12.34, Risk Solver Platform ultimately found that a reorder point of 35 and order quantity of 10. Because Risk Solver Platform is using a heuristic search algorithm, it might not find the same solution each time it solves a problem, and it might stop at a local (rather than global) optimal solution. Thus, on difficult problems, it is wise to run Risk Solver Platform several times to see if it can improve upon the solution it finds. Using a reorder point of 35 and order quantity of 10, 1000 replications were run resulting in an average service level of 98.2% and an average inventory of approximately 15.3 units per month.
12.15.5 Analyzing the Solution

Comparing the solution shown in Figure 12.34 to the original solution in Figure 12.26 we see that by using a reorder point of 35 and an order quantity of 10, MCC can simultaneously increase its average service level from 96.3% to 98.2% and reduce its average inventory level from approximately 26 units to 15 units. Another advantage of the optimal solution becomes apparent if we compare the behavior of the daily ending inventory balance under the original and optimal scenarios as shown in Figure 12.35.

In Figure 12.35, note that under the original policy (reorder point 28, order quantity 50) there are fairly wide swings in the amount of inventory MCC would be carrying for this product. Under the optimal policy (reorder point 35, order quantity 10), there is less volatility in the amount of inventory being held – which offers operational advantages from a warehousing and logistics perspective.
Creating Trend Charts

To create a trend chart like the ones in Figure 12.35 you add PsiOutput( ) functions to whatever cells you want to include in the chart (F6 through F35 in the MCC example). Then select Charts, Multiple Outputs, Trend on the Risk Solver Platform ribbon. In the resulting dialog, select the outputs you want to chart and click OK.

12.15.6 Other Measures of Risk

In the MCC example, Laura wanted to identify an inventory policy that would provide a 98% service level on average. While that might be a very reasonable goal, it would be wise to more carefully consider the downside risk associated with such a goal. Figure 12.36 displays the average service level distribution associated with the “optimal” solution to the MCC inventory problem.

Recall that Laura wanted a solution that provided an average service level of 98%. The mean of the distribution shown in Figure 12.36 is 98.2% and therefore satisfies Laura’s requirement. However, approximately 41.1% (or 411 out of 1000) of the trials in this simulation actually resulted in service levels that were less than 98%, with some as low as 89%. So if Laura uses a reorder point of 35 and an order quantity of 10, then in any month there is approximately a 41% chance that the actual service level will be below her desired average service level of 98%.

This discussion highlights the purpose of two other types of constraints available in Risk Solver Platform: the value at risk constraint and the conditional value at risk constraint. A value at risk (VaR) constraint allows you to specify the percentage of trials in a simulation that must satisfy a constraint. For example, Laura might want a solution where at least 90% of the trials have an average service level of at least 98%. (Clearly, the solution in Figure 12.36 violates such a VaR constraint.)

A VaR constraint only limits the percentage of trials that violate the constraint – counting a small violation the same as a large violation. In contrast, the conditional value at risk (CVaR) constraint places a bound on the average magnitude of the violations that may occur. Thus, the (CVaR) constraint is a more conservative version of the VaR constraint.
To illustrate the use of a VaR constraint, suppose that Laura would like there to be only a 10% chance of any particular trial’s average service level falling below 98%. This additional constraint and the resulting solution is summarized in Figure 12.37. Note that a “Chance” constraint was added to the model (in a manner very similar to how our other constraints were created). This constraint is of the VaR type and requires a 0.9 chance of the average service level (in cell M9) being at least 98%. This constraint will be satisfied if no more than 10% of the trials in a simulation have a service level less than 98%.

Re-running the optimization with this additional constraint resulted in a solution with a reorder point of 39 and an order quantity of 10. The frequency chart at the bottom of Figure 12.37 indicates that, as desired, approximately 10% of the simulation trials had average service levels of less than 98%.

12.16 A PROJECT SELECTION EXAMPLE

In chapter 6, we saw how Solver can be used in project selection problems in which the payoff for each project is assumed to be known with certainty. In many cases, a great deal of uncertainty exists with respect to the ultimate payoff that will be received if a particular project is undertaken. In these situations, Risk Solver Platform is a powerful aid deciding which project(s) to undertake. Consider the following example.

TRC Technologies has $2 million to invest in new R&D projects. The following table summarizes the initial cost, probability of success, and revenue potential for each of the projects.
TRC’s management wants to determine what set of projects should be selected.

### 12.16.1 A Spreadsheet Model

A spreadsheet model for this problem is shown in Figure 12.38 (and the file Fig12-38.xls). Cells C6 through C13 in this spreadsheet indicate which projects will be selected. Using Risk Solver Platform, we can define these cells to be decision variables that must take on discrete values between zero and one – or operate as binary variables. The values shown in cells C6 through C13 were assigned arbitrarily. We will use Risk Solver Platform to determine the optimal values for these variables.

In cell D14, we compute the total initial investment require by the selected projects as follows:

\[
\text{Formula for cell D14: } =\text{SUMPRODUCT(D6:D13,C6:C13)}
\]

In cell D16, we calculate the amount of unused or surplus investment funds. Using Risk Solver Platform, we can place a lower bound constraint of zero on the value of this cell to ensure that the projects selected do not require more than $2 million in initial investment funds.

\[
\text{Formula for cell D16: } =D15-D14
\]

A project has the potential to be successful only if it is selected. The success or failure of each project may be modeled using a binomial random variable using a single trial and the probability of success given in column E. Thus, we model the potential success of selected projects in column F as follows:

\[
\text{Formula for cell F6: } =\text{IF(C6=1,PsiBinomial(1,E6),0)}
\]

(Copy to cells F7 through F13.)

If a project is selected and successful, there is uncertainty about the revenue that it will generate. Because we have estimates of the minimum, most likely, and maximum possible revenue for each project, we will model the revenues for selected, successful projects using a triangular distribution. This is accomplished in column J as follows:

\[
\text{Formula for cell J6: } =\text{IF(F6=1,PsiTriangular(G6,H6,I6),0)}
\]

(Copy to cells J7 through J13.)

The profit associated with each project is computed in column K as follows:

\[
\text{Formula for cell K6: } =J6-C6*D6
\]

(Copy to cells K7 through K13.)

Cell K14 computes the total profit for each replication of the model. We will define this as an output cell using a PsiOutput( ) function.
Chapter 12 Introduction to Simulation Using Risk Solver Platform

47

Formula for cell K14:  =SUM(K6:K13)+PsiOutput( )

Finally, cell K16 computes the average (or expected) simulated total profit associated with cell K14. We will attempt to find the set of projects that maximize this value using Risk Solver Platform.

Formula for cell K16:  = PsiMean(K14)

Spreadsheet model for TRC Technologies' project selection problem.

Figure 12.38

12.16.2 Solving and Analyzing the Problem with Risk Solver Platform

The Risk Solver Platform settings used to solve this problem are shown in Figure 12.39 along with the best solution found and some additional statistics describing this solution. Risk Solver Platform identified solution that involves selecting projects 1, 2, 4, 6, 7, and 8, requiring an initial investment of $1.85 million and resulting in an expected profit of approximately $1.492 million.

The frequency chart in Figure 12.39 shows the distribution of possible profit values that might occur if TRC adopts this solution. Although the expected (mean) profit associated with this solution is approximately $1.49 million, the range of the possible outcomes is fairly wide at approximately $5.4 million (computed in cell K20 via =PsiRange(K14)). The worst-case outcome observed with this solution resulted in approximately a $1.6 million loss (computed in cell K18 via =PsiMin(K14)), whereas the best-case outcome resulted in approximately a $3.8 million profit (computed in cell K19 via =PsiMax(K14)).

Also in Figure 12.39, we see in cell K22 (labeled P(<$0) )that there is about a 0.089 probability of losing money if this solution is implemented. This probability was computed using the PsiTarget( ) function as follows,

Formula for cell K22:  = PsiTarget(K14,0)
In general, the \( \text{PsiTarget}(a, b) \) function returns the cumulative probability of distribution “\( a \)” taking on a value less than or equal to “\( b \)”. Thus, the formula in cell K22 computes the probability of the profit distribution in cell K14 taking on a value of less than \$0\). (Recall that all monetary values in this example are in \$1,000s.)

Similarly, as shown in cell K23, there is about a .367 probability of making less than \$1 million\) (or roughly a 63\% chance of making more than \$1 million\). Thus, there are significant risks associated with this solution that are not apparent if one simply looks at its expected profit level of \$1.49 million.\)

**Key Cell Formulas**

<table>
<thead>
<tr>
<th>Cell</th>
<th>Formula</th>
<th>Copied to</th>
</tr>
</thead>
<tbody>
<tr>
<td>K18</td>
<td>( \text{PsiMin}(K14) )</td>
<td>--</td>
</tr>
<tr>
<td>K19</td>
<td>( \text{PsiMax}(K14) )</td>
<td>--</td>
</tr>
<tr>
<td>K20</td>
<td>( \text{PsiRange}(K14) )</td>
<td>--</td>
</tr>
<tr>
<td>K22</td>
<td>( \text{PsiTarget}(K14,0) )</td>
<td>--</td>
</tr>
<tr>
<td>K23</td>
<td>( \text{PsiTarget}(K14,1000) )</td>
<td>--</td>
</tr>
</tbody>
</table>

**Settings and solution for maximizing average profit.**

**Figure 12.39**

### 12.16.3 Considering Another Solution

Because each of the projects is a one-time occurrence that can either succeed or fail, the decision makers in this problem do not have the luxury of repeatedly selecting this set of projects over and over and realizing the average profit level of \$1.49 million over time. As an alternative objective, TRC’s management might want to find a solution that minimizes the probability of having outcomes with profits
below $1 million (or equivalently, maximizing the probability of having an outcome of $1 million or more).

To pursue this new objective, we can simply optimize the model again with an objective on minimizing the value of cell K23. The solution to this problem is shown in Figure 12.40.

![Settings and solution for minimizing the probability of outcomes below $1 million.](image)

In Figure 12.40, notice that the expected (mean) profit for this solution is about $1.46 million, representing a decrease of approximately $33,000 from the earlier solution. The range of possible outcomes has also decreased to about $4.7 million, with a worst-case outcome of a $1.8 million loss, and a best-case outcome of almost $2.9 million profit. This solution reduces the chances of realizing a loss to approximately 9.6% and increases the chances of making at least $1 million to almost 73%. Thus, although the best possible outcome realized under this solution ($2.9 million) is not as large as that of the earlier solution ($3.8 million), it reduces the downside risk in the problem and makes it more likely for the company to earn at least $1 million – but it also requires a larger initial investment. It is also interesting to note that the probability of all the selected projects being successful under this solution is 0.2116 (i.e., $0.2116 = 0.9 \times 0.7 \times 0.8 \times 0.6 \times 0.7$), whereas the probability of all selected projects being successful under the first solution is only 0.0953 (i.e., $0.0953 = 0.9 \times 0.7 \times 0.4 \times 0.6 \times 0.9$).

So, what is the best solution to this problem? It depends on the risk attitudes and preferences of the decision makers at TRC. However, the simulation techniques we have described clearly provide valuable insights into the risks associated with various solutions.
12.17 A PORTFOLIO OPTIMIZATION EXAMPLE

In chapter 8, we saw how Solver can be used to analyze potential tradeoffs between risk and return for a given set of stocks using the idea of an **efficient frontier**. The efficient frontier represents the highest level of return a portfolio can achieve for any given level of risk. While portfolio optimization and efficient frontier analysis is most commonly associated with financial instruments such as stocks and bonds, it can be applied to physical assets as well. This will be illustrated using Risk Solver Platform with the following example.

In recent years, a fundamental shift occurred in power plant asset ownership. Traditionally, a single regulated utility would own a given power plant. Today, more and more power plants are owned by merchant generators that provide power to a competitive wholesale marketplace. This makes it possible for an investor to buy, for example, 10% of ten different generating assets rather than 100% of a single power plant. As a result, non-traditional power plant owners have emerged in the form of investment groups, private equity funds, and energy hedge funds.

The McDaniel Group is a private investment company in Richmond, VA that currently has a total of $1 billion that it wants to invest in power generation assets. Five different types of investments are possible: natural gas, oil, coal, nuclear, and wind powered plants. The following table summarizes the megawatts (MW) of generation capacity that can be purchased per each $1 million investment in the various types of power plants.

<table>
<thead>
<tr>
<th>Fuel Type</th>
<th>Gas</th>
<th>Coal</th>
<th>Oil</th>
<th>Nuclear</th>
<th>Wind</th>
</tr>
</thead>
<tbody>
<tr>
<td>MWs</td>
<td>2.0</td>
<td>1.2</td>
<td>3.5</td>
<td>1.0</td>
<td>0.5</td>
</tr>
</tbody>
</table>

The return on each type of investment varies randomly and is determined primarily by fluctuations in fuel prices and the spot price (or current market value) of electricity. Assume the McDaniel Group analyzed historical data to determine that the return per MW produced by each type of plant can be modeled as normally distributed random variables with the following means and standard deviations.

<table>
<thead>
<tr>
<th>Normal distribution return parameters by fuel type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fuel Type</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Std Dev</td>
</tr>
</tbody>
</table>

Additionally, while analyzing the historical data on operating costs, it was observed that many of the returns are correlated. For example, when the returns from plants fueled by natural gas are high (due to low gas prices), returns from plants fueled by coal and oil tend to be low. So there is a negative correlation between the returns from gas plants and the returns from coal and oil plants. The following table summarizes all the pairwise correlations between the returns from different types of power plants.

<table>
<thead>
<tr>
<th>Correlations between returns by fuel type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fuel Type</td>
</tr>
<tr>
<td>Gas</td>
</tr>
<tr>
<td>Coal</td>
</tr>
<tr>
<td>Oil</td>
</tr>
<tr>
<td>Nuclear</td>
</tr>
<tr>
<td>Wind</td>
</tr>
</tbody>
</table>

The McDaniel Group would like to determine the efficient frontier for its investment options in power generation assets.
12.17.1 A Spreadsheet Model

A spreadsheet model for this problem is shown in Figure 12.41 (and the file Fig12-40.xls on your data disk). Cells D5 through D9 in this spreadsheet indicate how much money (in millions) will be invested in each type of generation asset. The values shown in cells D5 though D19 were assigned arbitrarily. Notice in the Solve Options and Model Specifications pane that we have defined these cells to be decision variables that must take on values between zero and $1,000. We have also created a constraint that requires the sum of these values (computed in cell D10) to equal $1,000 (or $1 billion).

Formula for cell D10: \( \text{SUM}(D5:D9) \)
In column E we compute the number of MW of generation capacity purchased in each asset category as follows:

Formula for cell E5: =C5*D5
(Copy to cells E6 through E9.)

The cells representing random returns for each asset category are implemented in column F. Recall that we are assuming that correlations exist between these returns. Risk Solver Platform offers a number of different ways of dealing with correlations among variables. In this case, we model the correlations by including an appropriate PsiCorrMatrix( ) function as a third argument in the PsiNormal( ) function as shown below investments in gas fueled plants in cell F5.

Formula for cell F5: =PsiNormal(G5,H5,PsiCorrMatrix($C$14:$G$18,A5))
(Copy to cells F6 through F9.)

Note that the PsiCorrMatrix( ) function requires a (rank order) correlation matrix (C14 through G18 in our example) and an integer indicating which column (or row) in the matrix corresponds to the random variable being sampled (the value 1 in cell A5 in this example).

A Comment on Correlation

Any correlation matrix used in a Risk Solver Platform simulation must exhibit the mathematical property of being positive definite. The details of this property are beyond the scope of this book; however, it has to do with ensuring that the correlations are internally consistent with one another. For instance, if variables A and B have a high positive correlation and variables B and C have a high positive correlation, then variables A and C should have a fairly high positive correlation. The “Correlations” icon on the Risk Solver Platform ribbon offers a tool for checking if a correlation matrix is positive definite.

Also, it is important that, statistically speaking, correlation measures the strength of linear relationship between two variables. Sometimes variables are related in a nonlinear fashion. These nonlinear relationships cannot be summarized conveniently in a correlation matrix. Risk Solver Platform supports the modeling of nonlinear relationships between variables using its PsiSip( ) and PsiSlurp( ) functions that are described in the Risk Solver Platform user guide.

In cell F10, we calculate the weighted average return on the chosen investments in generating assets. This will also be the output cell that drives much of our analysis in this problem.

Formula for cell F10: =SUMPRODUCT(F5:F9,E5:E9)/SUM(E5:E9)+PsiOutput( )

In cells F21 and F22 we compute, respectively, the mean and standard deviation of the weighted average return in F10 for each simulation that is performed.

Formula for cell F21: =PsiMean(F10)
Formula for cell F22: =PsiStdDev(F10)

12.17.2 Solving the Problem with Risk Solver Platform

Recall that the McDaniel Group is interested in examining solutions on the efficient frontier of its possible investment options for these power generation assets. This requires determining the portfolios that provide the maximum expected (or average) return at a variety of different risk levels. In this case, we will define risk to be the standard deviation of a portfolio's weighted average return. Thus, in the
Solver Options and Model Specifications pane in Figure 12.41 we indicate that our objective is to maximize the mean value of the weighted average return calculated in cell F21 in our spreadsheet.

We also specify a variable requirement on the allowable upper bound of the standard deviation of the weighted average return (in cell F22). To do this, in cells F23 we use a PsiOptParam() function to identify a range of risk levels we would like to use in constructing an efficient frontier for this problem.

Formula for cell F23: =PsiOptParam(0.02,0.12)

In Figure 12.41, note that we also have defined a constraint requiring cell F22 (the standard deviation of the weighted average return) to be less than or equal to the value in cell F23. The PsiOptParam() function specifies a parameter that will be varied as multiple optimizations are performed. The number of optimizations to be performed is indicated on the Risk Solver Platform Options dialog shown in Figure 12.42. (This dialog is displayed by clicking the Options icon on the Risk Solver Platform ribbon.) When the solve icon is pressed, Risk Solver Platform now performs six optimization runs, automatically varying the value in cell F23 to six different values equally spaced between 2% and 12%.

![Specifying the number of optimizations to perform.](image)

Figure 12.42

Figure 12.43 displays a chart summarizing the maximum weighted average return found for each of the six optimizations. This chart corresponds to the efficient frontier for the McDaniel Group’s asset investment decision, summarizing the six portfolios it found and their relative trade-offs in terms of risk and return. The expected returns on these portfolios vary from 11.17% to 16.0% with standard deviations varying from 2% to 12% with higher expected returns being associated with higher levels of risk. Which portfolio is optimal for the McDaniel Group depends on the firm’s preferences for risk versus return. But
this analysis should help the firm select a portfolio that provides the maximum return for the desired level of risk – or the minimum level of risk for the desired level of return. Any of the six solutions can be inspected in detail on the spreadsheet by selecting the appropriate optimization from the “Opt #” dropdown circled in Figure 12.43. Risk Solver Platform’s PsiOptValue() function can also be used to retrieve specific values of interest from the various optimization runs.

12.18 SUMMARY

This chapter introduced the concept of risk analysis and simulation. Many of the input cells in a spreadsheet represent random variables whose values cannot be determined with certainty. Any uncertainty in the input cells flows through the spreadsheet model to create a related uncertainty in the value of the output cell(s). Decisions made on the basis of these uncertain values involve some degree of risk.

Various methods of risk analysis are available, including best-case/worst-case analysis, what-if analysis, and simulation. Of these three methods, simulation is the only technique that provides hard evidence (facts and figures) that can be used objectively in making decisions. This chapter introduced the use of the Risk Solver Platform add-in to perform spreadsheet simulation and optimization. To simulate a model, RNGs are used to select representative values for each uncertain independent variable in the model. This process is repeated over and over to generate a sample of representative values for the dependent variable(s) in the model. The variability and distribution of the sample values for the dependent variable(s) can then be analyzed to gain insight into the possible outcomes that might occur. We also illustrated the use of Risk Solver Platform in determining the optimal value of controllable parameters or decision variables in simulation models.
12.19 REFERENCES


THE WORLD OF MANAGEMENT SCIENCE

The U.S. Postal Service Moves to the Fast Lane

Mail flows into the U.S. Postal Service at the rate of 500 million pieces per day, and it comes in many forms. There are standard-sized letters with 9-digit ZIP codes (with or without imprinted bar codes), 5-digit ZIP codes, typed addresses that can be read by optical character readers, handwritten addresses that are barely decipherable, Christmas cards in red envelopes addressed in red ink, and so on. The enormous task of sorting all these pieces at the sending post office and at the destination has caused postal management to consider and adopt many new forms of technology. These include operator-assisted mechanized sorters, optical character readers (last-line and multiple-line), and bar code sorters. Implementation of new technology brings with it associated policy decisions, such as rate discounts for bar coding by the customer, finer sorting at the origin, and so on.

A simulation model called META (model for evaluating technology alternatives) assists management in evaluating new technologies, configurations, and operating plans. Using distributions based on experience or projections of the effects of new policies, META simulates a random stream of mail of different types; routes the mail through the system configuration being tested; and prints reports detailing total pieces handled, capacity utilization, work hours required, space requirements, and cost.

META has been used on several projects associated with the Postal Service corporate automation plan. These include facilities planning, benefits of alternative sorting plans, justification of efforts to enhance address readability, planning studies for reducing the time carriers spend sorting vs. delivering, and identification of mail types that offer the greatest potential for cost savings.

According to the Associate Postmaster General, “... META became the vehicle to help steer our organization on an entirely new course at a speed we had never before experienced.”


QUESTIONS AND PROBLEMS

Please see your textbook for the questions, problems, and cases for chapter 12.