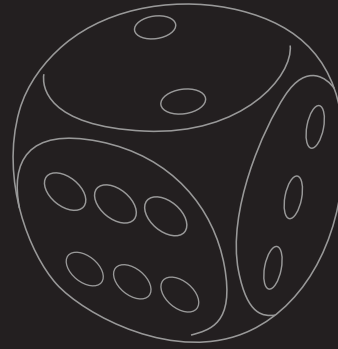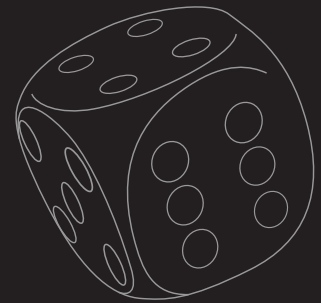Full Monte™ — The better approach to Schedule Risk Analysis.

Full Monte™

BARBECANA

FOREWORD

While Microsoft Project® and other project management systems give users single-point estimates of when projects will be complete, these deterministic estimates are often misleading. In fact, the only sure thing about a deterministic approach to scheduling is that the results will be proven wrong.

Quantitative assessment of risk is a much more robust approach to schedule management. Barbecana's Full Monte™, an add-in for Microsoft Project, addresses the issue of cost and schedule risk analysis by replacing single-point estimates with probability distributions and using Monte Carlo simulation to provide a much more realistic assessment of likely project outcomes.

While Full Monte offers powerful functionality, individuals with various levels of experience using Microsoft Project will find this fully integrated add-in user-friendly. Depending on your background and particular needs, you may begin learning Full Monte at any one of the following points:

- **User Guide, Chapter 1, Introduction:** Users without a background in statistics should read this review of risk analysis, the basic principles of statistics, and their application in Full Monte. After reviewing Chapter 1, refer to Chapter 2 for instructions on installing and activating Full Monte. The remainder of the User Guide offers a detailed reference to the functionality and capabilities of Full Monte, while the Guided Tour offers a quick introduction to the main features of Full Monte.

- **Guided Tour:** Users familiar with the principles of schedule risk analysis and who want an easy introduction to Full Monte's features may choose to start with Chapter 2 of the User Guide (which explains how to install Full Monte) and then move on to the Guided Tour, included with your download of the program. Having gone through the Guided Tour, the User Guide serves as a detailed reference on Full Monte.

- **User Guide, Chapter 2, Getting Started:** Users familiar with the principles of schedule risk analysis and who want to begin using Full Monte immediately may choose to start with Chapter 2 of the User Guide (which explains how to install Full Monte) and get started on their own projects, using the remainder of the User Guide as a detailed reference on Full Monte.

Whatever approach you take to learning Full Monte, we hope that you will find it a powerful tool and enjoy using it as much as we have enjoyed creating it. Comments, suggestions, and feedback are always welcome, so please feel free to contact us. Thank you for choosing Full Monte to help you better manage your projects.

Tony Welsh
TWelsh@Barbecana.com
713-595-6687
www.Barbecana.com

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: Introduction

Full Monte augments Microsoft Project (2007 and 2010) with important functionality, using Monte Carlo simulation to analyze the effects of uncertainty associated with a planned project. In this chapter we are going to explore:

1. The reasons why risk analysis is important in project planning.

2. The mathematical concepts of probability, probability distributions, and correlation.

3. How to specify data for Full Monte.

4. How to interpret results from Full Monte.

5. How Monte Carlo simulation works.

    NOTE: If you are familiar with these concepts and ready to get started with Full Monte, you may skip this chapter and move on to the Full Monte Guided Tour, which is supplied as a separate document when you downloaded the software, or begin with Chapter 2, Getting Started.

The importance of risk analysis

All management decisions are concerned with the future, and everything about the future is to a greater or lesser extent uncertain. Nowhere is this more true than in project planning, where ignoring uncertainty can have serious consequences on the quality of our plans. In fact, the only thing certain about a deterministic plan is that it will be wrong.

The basic data for a project plan normally comprises estimates of the durations and costs of each individual task; however, these cannot be precisely known in advance. The degree of uncertainty will depend upon the nature of the task and how closely it compares with previous tasks performed by the organization, but there will always be some degree of uncertainty.

Uncertainty about these individual task durations and costs results in uncertainty in projections of the total project duration and cost. Ignoring this uncertainty — that is, using single-point estimates — results in two types of errors:

1. At best, the deterministic value obtained for the project completion date or cost may approximate the middle of the range of possible outcomes. This means that there is roughly a 50% chance of achieving the targeted value. A prudent manager should require a plan that has a much higher chance of success.

2. Often the situation is much worse. The chance of achieving these deterministic values may be much smaller than 50% due to distributions being skewed (so that the expected value is greater

than the value used in a deterministic analysis) and to a phenomenon known as merge bias, to be explained at the end of this chapter.

Full Monte uses the technique of Monte Carlo simulation to address these shortcomings. Before we discuss how Monte Carlo simulation works, we need to understand some mathematical concepts related to probability and probability distributions.

Probability distributions have two roles in Full Monte and, indeed, in any Monte Carlo simulation:

1. They represent the inputs to the simulation in the current case task durations; and

2. They represent the results of the simulation in the current case dates and costs.

This is where the similarity ends. Input distributions are generally determined subjectively, though they may be based on past history. They are specified to Full Monte in terms of a choice of distribution type (such as an algebraic function defining the relative probability of various values) and two or more parameters for that function.

The output distributions, however, are empirical distributions that cannot be described in terms of an explicit function. It is as if we had a black box that generates simulated observations, and we count the number of observations in each of a number of ranges to build up a histogram.

So we need to understand probability in order to be able to choose appropriate distributions and interpret the results.

> *NOTE: It is beyond the scope of this user guide to give more than a cursory introduction to probability theory. The interested reader is advised to read one of the books on the subject listed in Appendix F.*

Random variables and probability distributions

The concept of a random variable is normally conceived in terms of the result of a repeatable experiment such as tossing a coin or throwing a die. The result of any one such experiment, or "trial," cannot be predicted, but the behavior over a large number of trials conforms to a pattern. In the case of throwing a die, we know that if we do a large number of trials, each possible value from 1 through 6 will come up roughly one-sixth of the time. Therefore, we postulate that as the number of trials approaches infinity, the proportion of times that each number comes up will approach exactly one-sixth. We say that the probability of each possible result is one-sixth.

The probability distribution of a random variable defines the relative probabilities of that variable taking any of its possible values. It is expressed as either the frequency function (normally denoted by lower case f) or the cumulative distribution function (denoted by upper case F):

f(x) denotes the probability of the variable being x; while

F(x) denotes the probability of the variable being less than or equal to x.

In the above example of the die throw,

f(x) = 1/6; and

F(x) = x/6

for integer x such that 1 <= x <= 6.

This is an example of a discrete distribution, because x can take only a finite number of discrete values. Most distributions are continuous distributions, called that because the possible values of x are continuously variable, like, for example, the daily maximum temperature. For continuous distributions, the frequency function does not represent actual probabilities because the probability of a continuous variable being exactly equal to any particular value is zero. Rather it represents the probability density, so that the probability of the variable being within a small range of x, say δx, is f(x) times δx. (Readers not comfortable with calculus need not worry about this distinction.)

NOTE: Because we can measure any real variable only to a certain degree of accuracy, one could argue that all distributions are discrete. However, it is more convenient to regard them as continuous. Full Monte uses continuous distributions even though it then rounds the results to the nearest minute for convenience in storage and processing.

The cumulative distribution function (CDF) is more consistent in its applicability to both discrete and continuous distributions because it always represents an actual probability and is thus perhaps easier to understand. By definition, its value always ranges from 0 (for sufficiently low values of x) to 1 (for sufficiently high values of x). For continuous distributions this function normally has an S-shape, whereas for discrete distributions it will comprise a step at each discrete allowable value. We will meet the CDF again when we talk about the interpretation of Full Monte results later in this chapter.

When defining the input data for Full Monte, we use the frequency function because it gives us a better picture of what the distribution looks like. For example, the most likely value is clearly represented by the highest value of the frequency function, whereas all CDF S-curves tend to look alike and to determine the most likely value one would have to find the steepest gradient.

Subjective probability

It is important to note that the concepts of probability and probability distributions apply even when there is no opportunity to do repeated trials. Faced with a single throw of a die, a gambler will still bet on the assumption that the above frequency function applies.

Perhaps even more pertinent to the case at hand, a gambler will bet on the outcome of events. For example, the winner of a horse race where there is no clear model of the probability distribution. Whether or not the gambler takes a particular bet will depend upon his subjective estimate of the probability of a horse winning, even though there is no history of having done exactly that race — at the same venue, with the same field of competitors (all at their current ages and in their current state of fitness), under the same weather conditions, etc. The "Names" at Lloyd's of London make similar judgments when they insure a ship.

There is therefore a sense in which probability exists even in the absence of the possibility of multiple trials. This subjective probability is important to us, because the project manager is in a similar situation to the horse fancier or the Lloyd's Name. He will not have done the exact same project even once before, let alone enough times to have built up an approximation to the true probabilities involved.

Furthermore, different people will make different estimates of these probabilities, which leads us to an important recognition: The probability assigned to a future event is not in general a fundamental property of that event, but rather it is a property of the state of knowledge of the estimator. For example, someone waiting for a taxi might estimate from previous experience that the chance of one coming by in the next 30 seconds is 20%. A taxi driver around the corner, however, might say it was close to 100% — one event, two states of knowledge, two different and equally valid estimates of probability.

> *NOTE: It could be argued that there is no such thing as a truly random variable. Even in the case of throwing a die, it is only lack of knowledge that makes it seem random. That is, if we knew enough about the circumstances of the throw, the aerodynamics of the die, and so forth, we could be certain about the outcome. This philosophical point need not concern us here, except to note that probability is in our minds.*

One objection often raised in Monte Carlo simulation is the fact that the data is subjective: "How can I estimate the probability distribution when I've never done this before?" There are several answers to this:

1. The fact that you have not done a task before and therefore know little about its likely duration is part of the uncertainty we are modeling. This uncertainty is not a feature of the task but of the state of knowledge of the estimator, so his estimate should span a wide range to reflect this lack

of knowledge. (Of course, if possible you should obtain your estimates from the person whose experience makes him the best able to make them.)

2. If you don't know enough to specify a range, which can be as wide as you feel necessary, how can you possibly know enough to produce a single-point estimate, which is almost bound to be at least somewhat wrong?

3. What choice do you have? The future is uncertain and the task is in the future, so we must do our best. In much the same way that keeping one's money under the mattress is an investment decision, using single-point estimates for task durations amounts to asserting that there is no uncertainty about them. (Otherwise known as "rowing with the oars we have.")

Choosing probability distributions for task durations in Full Monte

Full Monte supports five distribution types, which are discussed in detail in Appendix A:

1. Beta

2. Lognormal

3. Normal

4. Triangular

5. Uniform

We have deliberately kept this list short in the interest of simplicity and because in the final analysis we have found that the exact shape of the distribution is less important than the parameters. As Sam Savage says in his excellent book, The Flaw of Averages, "In the land of averages the man with the wrong distribution is king."

Nevertheless, there are five choices. How do you pick the right one? As noted above, the choice is subjective, so there is no single right answer. Nevertheless, we can offer some guidance. Rather than treat the choices in the above (alphabetical) order, here are some observations.

The **uniform distribution** is the simplest but generally the least realistic. It represents the case when all values between the two limits (optimistic and pessimistic) are equally likely. One example would be the time taken for a bus to arrive when you know they come every 10 minutes but you don't know the actual schedule. But for most task durations this is an unlikely choice, because in most cases durations are relatively less likely to be close to the ends of the specified range. Unless you have specific reasons to think the distribution is uniform, this choice should be avoided.

If you think the distribution is symmetrical — the outcome is as likely to be less than the most likely value as greater than it — then the **normal or lognormal distribution** is a good choice. The normal distribution is often used to model natural processes .

> *NOTE: One objection to the normal distribution is that in its true form, it has infinitely long tails. This means there is a small but finite probability of the outcome being very large or — worse still — negative. Full Monte avoids this by cutting the distribution off at the ends. The values entered as the optimistic and pessimistic values are interpreted by default as "3-sigma" values, which means that the area between them is just about 99.73% of the total area under the true normal curve.*

The **lognormal distribution** can be very similar to the normal distribution, except that the lognormal distribution has an absolute lower bound of zero. The lower bound entered by the user, however, is (like for the normal distribution) interpreted as a 3-sigma limit. The closer it is to zero, the more skewed the distribution. There is some empirical evidence to suggest that the lognormal distribution is the most appropriate distribution in most circumstances. (For empirical arguments for the use of the lognormal distribution, see the URL references in Appendix F.)

If you have reason to suppose that the distribution is skewed — and in a way that you wish to control — then you need to choose either the triangular or beta distributions. The choice is largely a matter of taste. The **triangular distribution** is easy to understand but does not appear often in nature (although, for example, the sum of two uniformly distributed variables has a triangular distribution).

The **beta distribution** has a smoothly curved shape that some find more plausible than the triangular, but — especially when severely skewed — it has a long, thin tail. (Though setting Full Monte's unique confidence interval parameter, discussed in the next section, to something less than 100 largely overcomes this problem.)

> *NOTE: The beta distribution as implemented in Full Monte is a subset of the true beta distribution, sometimes called PERT-beta because it was made popular in project management by its use in PERT. It is best regarded as just a conveniently shaped distribution that is close to normal when not skewed, but which can be skewed to an arbitrary degree.*

Whichever distribution type you choose, Full Monte gives you immediate visual feedback as to what the frequency function looks like. What matters more than the exact shape are the parameters you choose and, consequently, the range of possible values and the extent to which the height of the frequency function falls off toward these limits.

Setting the Distribution Parameters

Whichever distribution type you choose, you have to specify at least two parameters, namely an optimistic value and a pessimistic value. These values should normally represent your estimates of the best and worst possible outcomes having regard to all possible circumstances. (Again, setting Full Monte's unique confidence interval parameter, discussed in the next section, to something less than 100 largely overcomes this problem.)

In the case of the triangular and beta distributions, you also need to set the most likely value, which as it suggests is the value you estimate to be the most likely outcome. This should be between the optimistic and pessimistic values. If you specify a value that is equal to or close to either of these, the distribution will be highly skewed and, in the case of the beta, will have a very long, thin tail. This means that the probability of outcomes toward the other end of the range, the end that is not close to the most likely value, is very close to zero and in practice may never occur during the simulation. This will be apparent from the pictorial representation, and you might want to adjust the parameters to more accurately reflect your estimates.

Another parameter that defaults to 100% that you might wish to change is the confidence interval. Confidence interval is a measure of the confidence you have in these optimistic and pessimistic estimates. For example, if you are not comfortable setting absolute limits on the possible outcomes, you can set this to 90% and set the optimistic and pessimistic values in such a way that you expect just 90% of possible outcomes to fall between them.

This is also a way to overcome the problem of the long, thin tail on a heavily skewed beta distribution. Even setting the confidence interval to 99% instead of 100% will have the effect of pushing this tail well beyond the optimistic or pessimistic value you set.

Correlations

One other aspect of the input data for Full Monte is the ability to specify correlations between the durations of various tasks. Correlation occurs when two or more random variables are not independent. Let us consider what we mean by independent. When we throw two dice, we assume that the results are independent of one another. The chance of the second die coming up 6 is 1/6 regardless of the result of the first die.

As has been noted before, if we add the results of two independent dice throws together, we get a variable that has a triangular distribution, peaking at 7, the most likely value because there are more possible combinations of values that add up to 7 than any other. This case is represented by a correlation coefficient of 0.

If, on the other hand, we could somehow make the dice throws perfectly correlated, their values would always be the same and the distribution of the sum would be uniform with only even values 2, 4, etc. This is represented by a correlation coefficient of +1.

> *NOTE: It is also possible to have negative correlations. If the dice were perfectly negatively correlated with each other (a coefficient of -1), then every throw would produce the same sum, 7. (1 would always be paired with 6, 2 with 5, and 3 with 4.)*

More common variables will be somewhat but not perfectly correlated, represented by a correlation coefficient between 0 and (+ or -) 1. This means that if one of the values is higher than average, then there is a tendency for the other one to be higher (or in the case of a negative coefficient, lower) than average.

> *NOTE: This is necessarily a very simplified treatment of a quite complicated subject, and again the interested reader is referred to the bibliographical references in Appendix F.*

Correlations are treated in Full Monte by reference to what we call correlation sources that represent the reasons why we expect two or more tasks to be correlated. The advantages of this are explained in Chapter 5, Investigating Correlation Sensitivity. This might seem strange, since it is commonplace for statisticians to point out that correlations do not necessarily indicate causation. But this relates to empirical data in which correlations have been observed. In the case of subjective data for Full Monte, we are dealing with correlations that are supposed to exist in the estimator's model of the world. Given that he supposes them to exist, a priori, he presumably has a reason to suppose they exist, and this implies causality. If two tasks are correlated with one another, this is either because one affects the other or because they are both affected by the same outside factor, or correlation source.

Correlations might seem rather academic, but they are important because they can have a very significant effect on the project outcome. For example, consider a project with 10 tasks. Each has a cost that is normally distributed with a mean of $1,000 and a standard deviation of $100. The expected total cost is $10,000. But:

- If they are independent, the chance of the cost exceeding $11,000 (a 10% overrun) is less than 1 in 1,000, *while*

- If they are perfectly correlated, the chance of the same overrun is about 16%, *and*

- If the correlation coefficient between each pair of tasks is 0.5, the chance is about 10%.

> *NOTE: Correlation sources in Full Monte work quite differently from the "drivers" promoted by some vendors, though they seem superficially similar. This is explored in Appendix B, Correlations.*

How Monte Carlo simulation works

Monte Carlo simulation gets its name from the casinos that make the European city famous. Conceptually, Monte Carlo simulation is very simple. In order to find out the distribution of possible outcomes, we merely perform the process a large number of times. We build up a histogram from the results, representing the probability distribution from the individual outcomes of each trial.

Of course we cannot really perform a project many times. Indeed, we cannot even perform it once without defeating a principal purpose of risk analysis, namely to decide whether or not the project is feasible in the first place.

We can, however, pretend to do so, and this is exactly what Monte Carlo simulation does. Instead of doing time analysis once based on fixed duration lengths for each activity, we simulate the project several hundred or several thousand times. Each time, we use a different set of data for durations and costs sampled from the probability distributions of these uncertain values.

The sampling is done so that the probability of selecting a particular duration in the simulation is the same as our subjective estimate of the probability of that value actually occurring. With definite values for activity duration and costs thus supplied, we can calculate such things as project completion date and total cost for each trial, just as we would with deterministic data.

By doing this with thousands of samples, a picture is built up of the distribution of all the values of interest. These are stored in a format that can be displayed as histograms (approximating the frequency function) and S-curves (representing the cumulative distribution function [CDF]). For tabular reports, these distributions can also be represented by their means and standard deviations and by various percentiles.

During the course of the simulation, Full Monte also gathers data that enables it to create estimates of various measures of sensitivity of the early finish of the project (or of a preselected milestone) and of the cost of the project (or preselected summary task) to each of the input task durations and to each of the correlation sources. These form the starting point for a more sophisticated analysis of the important sensitivities as discussed later in this chapter.

Interpretation of Full Monte results

Most of the outputs from a Monte Carlo simulation are distributions, but unlike the input distributions we do not know the mathematical functions that define their shapes. What we have are actual samples from these distributions. Since we cannot define them in terms of their frequency functions, we instead display and summarize them in various ways:

- Histograms depict the frequency function graphically.

- S-curves depict the cumulative distribution function graphically.

- Percentiles represent points on the cumulative distribution in a form that can be displayed in tables.

- The expected value is a measure of the central tendency of the distribution.

- The standard deviation is a measure of the variability.

All these are calculated for the following values for the project as a whole and also for every task in the network:

- Early start and finish

- Late start and finish

- Free and total slack

- Cost

There are also some results that do not fit into the above framework. These are described in Appendix E, Full Monte Output Fields.

For the sake of specificity, we will concentrate on the early finish date for the project, which is probably the most important single statistic. Everything that follows applies equally to all the other values.

The most common way to view the distribution of a particular value like the project early finish date is to look at the histogram (showing the frequency function) and S-curve (showing the CDF). Any of the graphs can be viewed by first selecting the tabular report Graphs, and then clicking on the button representing the data of interest. Most useful is the S-curve, because generally one is more interested in completing the project *by* a certain date rather than *on* a certain date. The S-curve tells you the probability of completing the project by any date one chooses. (This can be read from the axes or by hovering over the curve.)

One can also view tabular reports, on which the results for all tasks can be presented. The complete distributions of the various data from each task cannot be shown in the table, though you can include columns of buttons that take you straight to the corresponding graphs. The distributions can be summarized in terms of their means and standard deviations and by various percentiles. For example, the 80th percentile, sometimes known as "P80," is the value that the actual value has an estimated 80% chance of not exceeding.

NOTE: If you produce a tabular report showing, for example, P80 values, do not expect them to add up. Some people apparently expect them to, but unless all of the values are perfectly

*correlated, they will not. This is because the chance of everything going wrong is generally
smaller than the chance of one thing going wrong.*

Interpreting sensitivity analysis results

This relates to the sensitivity of certain output results to particular input data. Full Monte supports
sensitivity to two types of input:

- The durations of individual tasks, *and*

- The effect of outside factors (correlation sources) that affect multiple tasks.

Again, in the interest of specificity, we will discuss the sensitivity of the project's early finish to various
input data, though all that follows applies equally to other cases.

The only way to get a complete and accurate picture of the sensitivity of the project finish date to a
particular input is to do two complete simulations. One simulation will be based on an optimistic value
for the input in question and one will be based on a pessimistic value, allowing all other inputs to vary as
specified. Full Monte provides the capability to do this, but it would be too time-consuming to do this
for every possible input. (For example, for a network with 500 tasks, we would need to do 1,000 full
simulations, each with 10,000 iterations, for a total of 10 million trials.) Full Monte therefore provides
several approximate measures of sensitivity that are quick to calculate and are accurate enough to
indicate which tasks are worth investigating in more detail.

For sensitivity to individual task durations, these measures are:

1. The sensitivity index, the criticality percentage multiplied by the standard deviation of the
   duration and divided by the standard deviation of the project early finish.

2. An estimate of the proportion of the variance in the project early finish that is attributable to
   variation in the current task's duration.

3. An estimate of the expected project early finish if the current task's duration realizes its most
   optimistic date.

4. An estimate of the expected project early finish if the current task's duration realizes its most
   pessimistic date.

These values are displayed on the tornado chart. The bars on this chart stretch between the dates
defined in points 3 and 4 above and are sorted in descending order of the length of these bars.

NOTE: These two dates are stored (and the bars appear) only if they are considered to be significantly different from one another. Doing more iterations may make more such differences statistically significant and will result in more bars being displayed.

For sensitivity to outside influences (correlation sources), Full Monte produces only item 2 above (an estimate of the proportion of the variance in the project early finish `that` is attributable to variation in the current task's duration). The results are viewed in a dialog box invoked by clicking on **correlation sensitivity** on the histogram displayed at the end of a simulation run. (It is necessary to check the box to request this report when running the simulation in order to have access to this sensitivity information.)

In both cases, it is possible to get the full analysis referred to above. In the case of the **tornado chart**, it is accessed by clicking on a bar. In the case of the **correlation source**, it is accessed by selecting one or more correlation sources from the dialog box referred to in the preceding paragraph. In either case, the result is a pair of cumulative distribution functions, one based on the optimistic assumption about the task or factor being investigated and one based on the pessimistic assumption. The Y-axis is marked not with the absolute values from the X-axis but with the differences between the two curves at that point. From this, one can tell how sensitive the whole distribution of the project early finish (or other result of interest) is to the task or factor.

NOTE: The above also applies to cost. The sensitivity measures are calculated using the same methodology and are therefore only as accurate as the sampling error allows. It may seem odd to do it this way because in simple cases the exact result can be determined in one's head. For example, if the cost of a task can vary by + or – $500, then sensitivity of the total project cost to this task will normally also be + or – $500. However, the potential existence of escalation means that a delay in one task may actually influence the cost of a successor task and hence the cost of the whole project in ways that cannot easily be predicted. Branching can also confuse the issue. For these reasons, Full Monte always uses the statistical approach rather than a shortcut.

Merge bias

Merge bias is probably the single biggest reason why deterministic schedules (and PERT schedules) are too optimistic and therefore why Monte Carlo simulation is so important. It results when two or more parallel paths merge either at the end of the project or as predecessors to some intermediate task or milestone and when two or more of these paths have a chance of being critical.

NOTE: Even if a schedule has multiple finish tasks, these all "merge" for the purpose of this section, in the sense that the project is not finished until every path is finished.

Because the successor cannot start until *all* its predecessors are complete, it is always dependent on the path which happens to finish last even though that path is not always the same one. This results in the

expected start date of the successor being later than the expected finish dates of any of its predecessors. This delay is known as merge bias.

Merge bias can be illustrated by analogy with the throwing of a pair of dice. Each die has equal chances of coming up with any value 1 through 6, so the expected value is 3.5. But if we look at the distribution of the greater of the two scores, which is equivalent to waiting for the longer of two predecessors, the expected value turns out to be about 4.5. (For more on the math behind this assertion, please see the bibliographical references in Appendix F.) This additional 1 represents the merge bias.

Merge bias builds up over the duration of the project, causing greater and greater discrepancies between a deterministic plan and reality, so that while early milestones may be met the later ones are increasingly likely to be missed. It is truly a bias toward optimism in deterministic schedules, in the sense that there are no countervailing tendencies that might work the other way.

CHAPTER 2: Getting Started

In this chapter, we'll explain how to install, activate, and register Full Monte. We will then discuss field mapping and migrating Risk+ data.

Installation

When downloading Full Monte from the Barbecana website, choose to **Save** it rather than run it. Then double click on the downloaded setup program and follow the instructions.

Activation

Full Monte is protected by a licensing system; however, you will receive a free 30-day trial after your first installation. The first time you enter Microsoft Project, you will be given the opportunity to activate Full Monte on a screen that resembles the following:
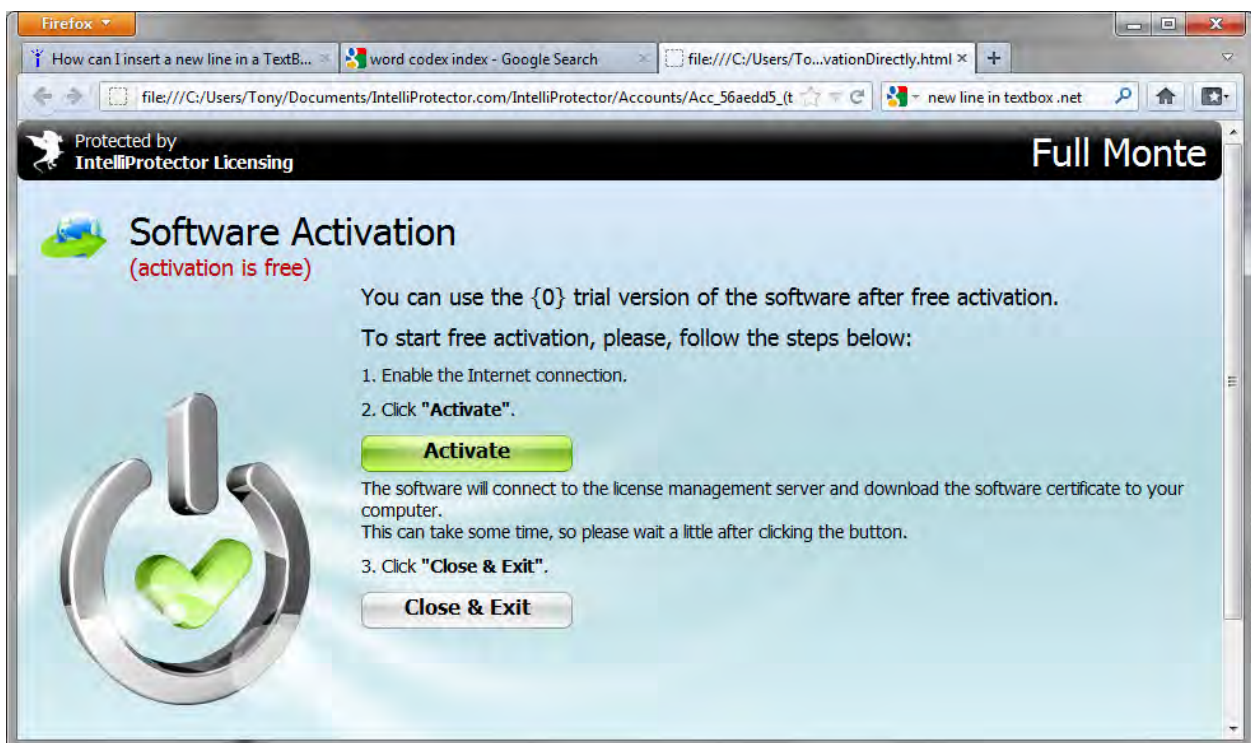

Figure 2.1

Activation is free but requires Internet access. (If you are in a secure environment without Internet access, you will get a slightly different screen enabling you to activate the software by email using a different computer.) To activate your free trial, click on the **Activate** button. Full Monte is now activated.

NOTES:

1. *If you click on __Close & Exit__, Project will close down. (This is a feature of the third-party licensing software and is outside the control of Full Monte.) When you reenter Project, you will not see the Software Activation dialog box again. The Full Monte menu items will appear but will not be activated. Later, when you are ready to use Full Monte, click on any of the Full Monte menu options to request another opportunity to activate Full Monte. (Full Monte shows the activation screen only on start-up in order to avoid the possibility of your losing data if it causes Microsoft Project to close down. After the first time, it does so only after you have requested it as described here.)*

2. *If you have downloaded and installed Schedule Inspector as well as Full Monte since the last time you launched Project, you will get the opportunity to activate only one of the two products. To activate the other one, exit Project and re-launch it after waiting 20 or 30 seconds.*

Registration

Registration gives you permanent access to Full Monte after you have purchased a license. Both the purchase and registration are done via the same dialog box, which looks like this:
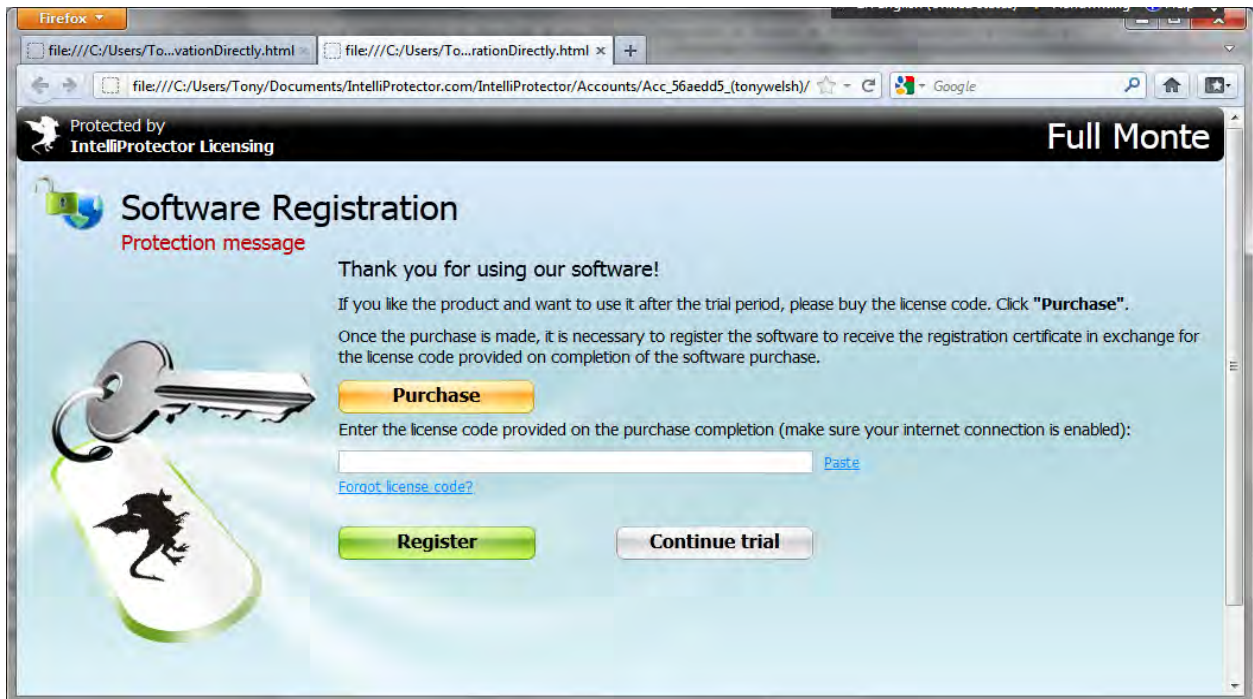

Figure 2.2

The dialog box appears periodically and with increasing frequency when your trial period is about to elapse. (Again, if you do not have Internet access, the screen is slightly different and you can register by email.) You can also access the software registration dialog box via the menu, which brings up the dialog shown in figure 2.2. We recommend purchasing Full Monte a few days before the trial period elapses to ensure uninterrupted use of the software.
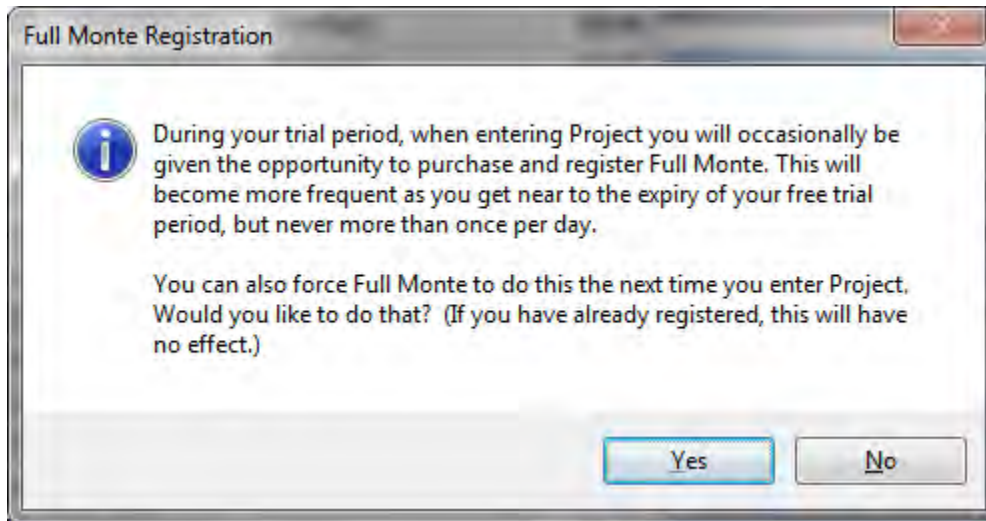

Figure 2.3

NOTES:

1. *After the trial has elapsed, if you have not yet registered the software it will be disabled. When you then try to use a Full Monte function, you will get a dialog box similar to that shown in figure 2.3. If you click on Yes, the dialog box in figure 2.2 will appear the next time Project is launched.*
2. *As with activation, if you need to register Schedule Inspector as well as Full Monte, you will need to launch Project twice with a 20 to 30 second delay between them.*

When you see the Software Registration window shown in figure 2.2, you have two options:

1. To *purchase* the software, click on the **Purchase** button. You may use a credit card or your PayPal account. When completed, you will receive a license key by email.

2. To *register* the software, copy the license key into the designated space in the Software Registration window shown in figure 2.2. Then click on the **Register** button.

If you press the **Continue Trial** button, you will be returned to Microsoft Project or to the Full Monte function you requested. (If your trial period has elapsed, the **Continue Trial** button will not appear.) Closing the dialog box without registration will shut down Project. You will not see the registration dialog box again until you request it through the menu, as described above.

Field mapping

Before you use Full Monte on live data, you should check the field mappings described in this chapter. You may need to modify the field mappings, especially if you have used Project with another add-in such as Risk+.

> *NOTE: If you simply want to take a look at a demonstration of Full Monte, work on an expendable copy of one of your projects or create new test data of your own. You may skip the rest of this chapter for now.*

Full Monte and other add-in products for Project make use of user fields supplied by Microsoft for that purpose. For example, let's say the data for the cost histograms is held in the user field Text30. We say that this data is "mapped" to Text30. To avoid collisions between two uses for the same user field, Full Monte allows you to specify any mapping to a field of the appropriate type, avoiding conflicts with any other add-in products you may use. You may also choose not to map some output fields that you are not interested in.

> *NOTE: Collisions will not be an issue if you have two add-ins using the same fields, but never use them on the same projects.*

The first thing you need to do when mapping fields is to find out which user fields are used by other add-ins or by users entering data directly. If you have been using Risk+, you can do this by selecting **Field Usage** from the drop-down menu in that product. You will then get a dialog box very similar to the Field Mapping dialog box in Full Monte.

To change mappings, go to the **Add-Ins** ribbon, then **Full Monte Administration**, then **Edit Field Mappings** from the drop-down menu. You will see the Field Mapping Information dialog box, as shown in figure 2.4.

> *NOTE: Here and throughout the remainder of this user guide, the instructions are written from the perspective of the Project 2010 user. If you are using Project 2007, there is no concept of ribbons. Instead, all the Full Monte options will appear on the Full Monte drop-down menu.*
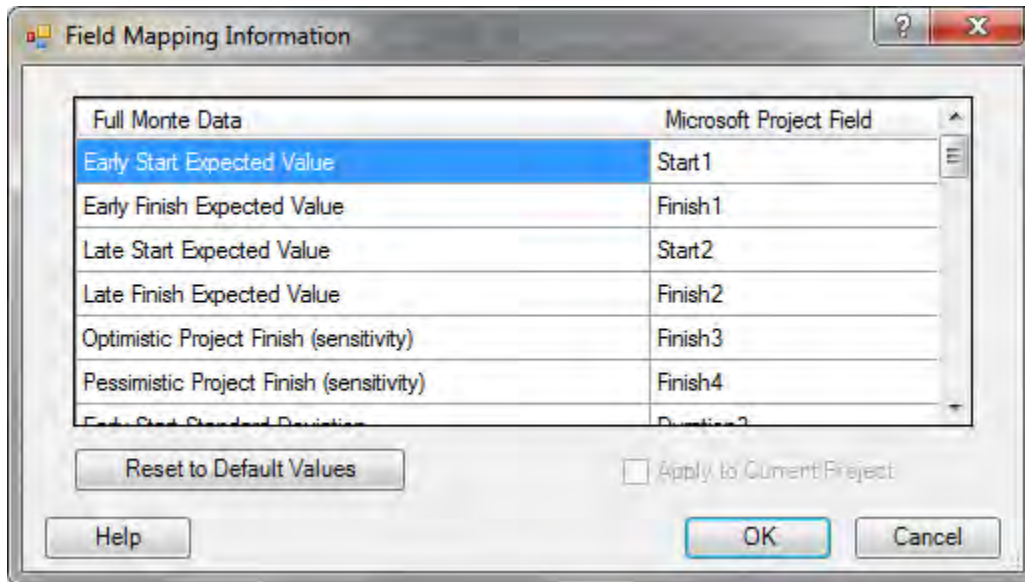
Figure 2.4

The first column is a description of the Full Monte Data. In the second column, you see the names of the user fields to which they are mapped. To change a mapping, click on the field in column 2 that you wish to change and select a new mapping from the drop-down list that appears (see figure 2.5). In most cases, the first option on that drop-down list will be **(None)**, meaning there is no mapping and the data will not be saved. Use this if you are short of free fields and this is an item in which you are less interested. For example, early finishes are generally of more interest than early starts or late dates.

NOTE: The one exception to this is that you must provide a mapping for Full Monte Input Data. *There would be no point in running Full Monte if you could not store the input data. Of course, there would also be no point if you did not provide a mapping for at least one item of output data.*
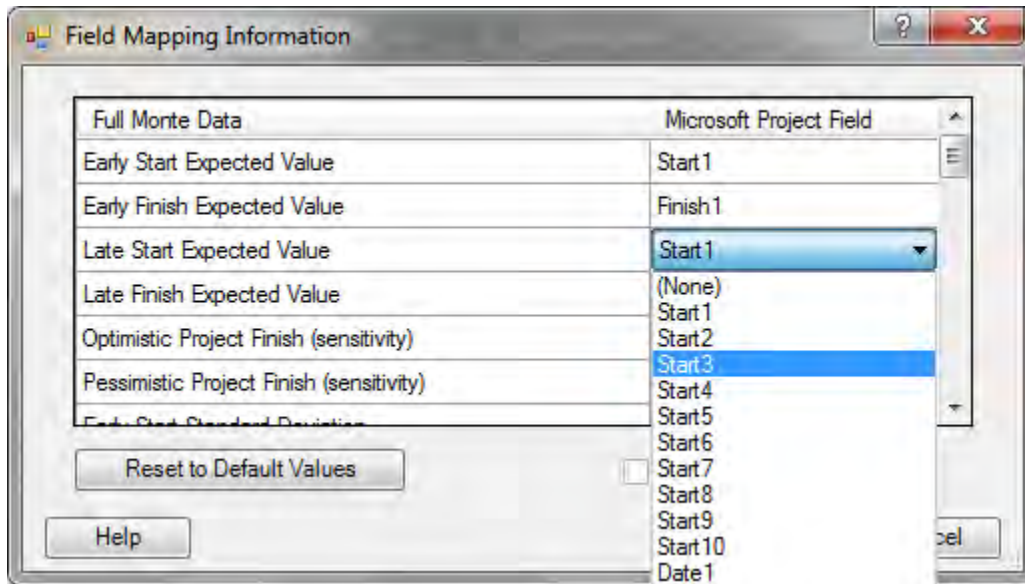
Figure 2.5

When you are finished and before you click on **OK**, there is a checkbox to **Apply to Current Project**. (This will be disabled if there is no project open or if the mappings already match.) If you do not check this box, you will change only the default mapping for projects new to Full Monte.

> NOTE: If you do not check this box, you will be warned the first time you open the project in a given session — as well as other projects with different mappings — that the mappings are different and given an opportunity to change them. Full Monte supports different mappings on different projects, but it is probably less confusing to be consistent, especially if you intend to use the external subproject feature in Project.

Migrating Risk+ data

If you have been using Risk+, you might want to start by migrating one of your projects.

> *NOTE: Before you migrate a project, you are strongly advised to copy it first, especially if you might go back to Risk+.*

The first thing you need to do is find out which fields Risk+ uses, as described in the preceding section, Field mapping. You will see a list of the mappings for a number of data items, but you need to take note of the mappings for just the first six (Full Monte migrates only the input data):

- • Minimum Remaining Duration

- • Most Likely Remaining Duration

- • Maximum Remaining Duration

- • Duration Distribution Curve

- • Task Reporting ID

- • Branch Definition

> *NOTE: You cannot migrate correlation information, because correlations are handled very differently in the two products.*

Now open the project you wish to migrate. (Remember to make a copy of the project first, especially if you think you may go back to using Risk+.) Go to the drop-down menu in Full Monte and select **Migrate Risk+ data**. You will see the dialog box shown in figure 2.6.

Figure 2.6

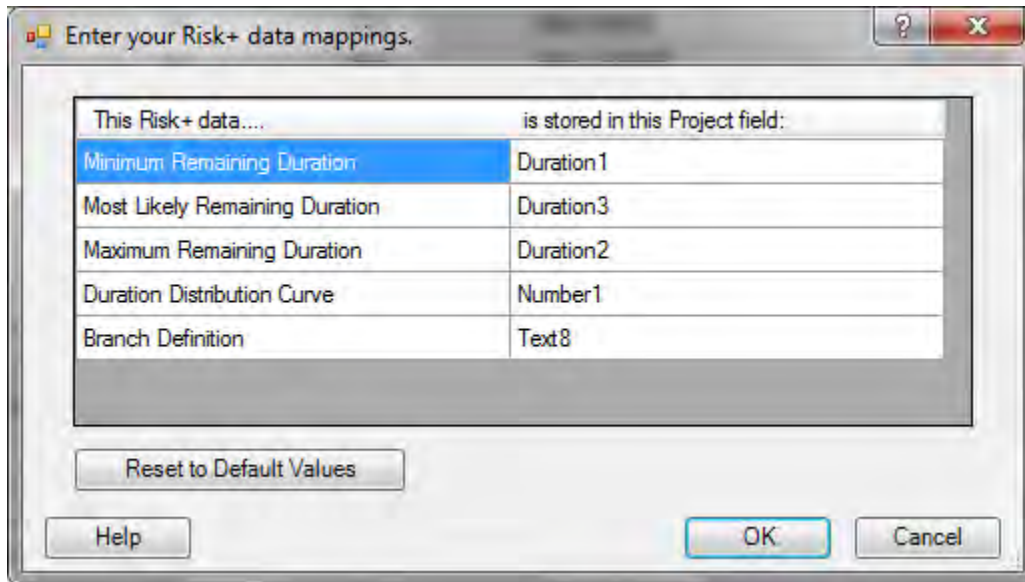Select each field in turn, clicking on the field name and using the drop-down menu that appears to specify the same mapping as you noted in Risk+. When you have finished, click on **OK**. The Risk+ data will be reformatted and moved to the appropriate Full Monte field.

Error messages

There are a couple of error messages you may encounter, Conflicts and Bad Data.
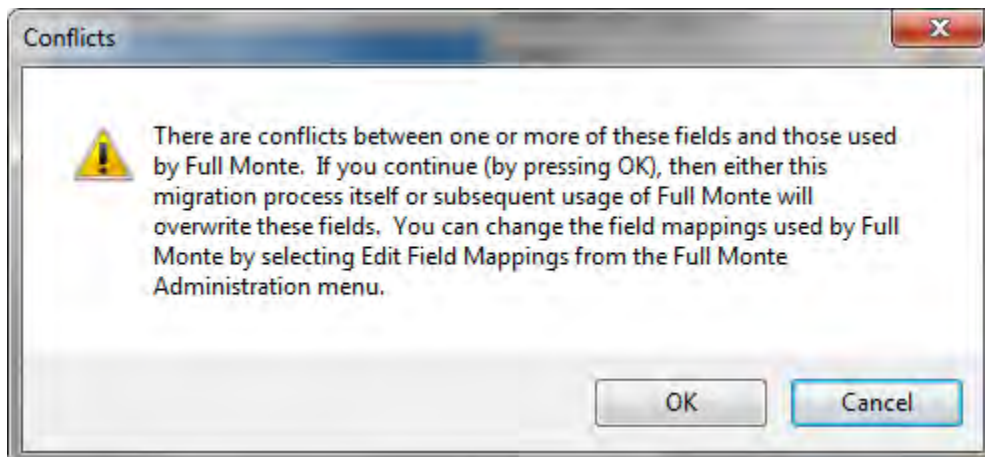
Conflicts is shown in figure 2.7.

Figure 2.7

If you are not worried about overwriting Risk+ fields, click on **OK**. Otherwise, press **Cancel** and go back to the field mapping dialog box in each product to spot the problems and revise the mappings. When the process is completed successfully, save the project.

> *NOTE: If you did not previously make a copy, this would be a good time to **Save As** to a new name.*

When you migrate the rest of your Risk+ projects, Full Monte will remember the mappings you entered here. You will not have to do anything on the dialog box except click on **OK**.

Bad Data is the second error message you may encounter, shown in figure 2.8.
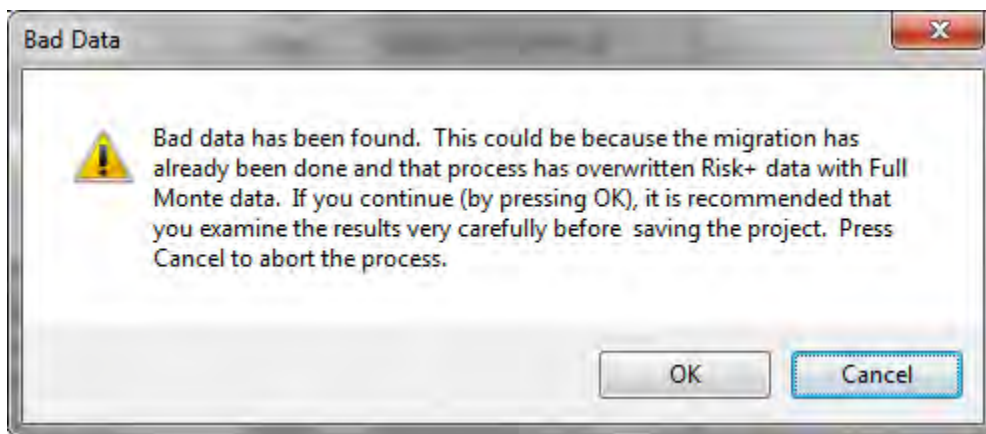
Figure 2.8

If you received this error message, it is because Full Monte was unable to interpret the Risk+ data. There are three possible reasons for this:

- You may have made a mistake in the mappings you entered.

- This project may have never actually been used in Risk+.

- This project may have already been migrated and used in Full Monte, which has overwritten one or more of the Risk+ fields.

If you continue, you still have the opportunity to examine the data before saving, preferably to a new project name.

CHAPTER 3: Data Entry

This chapter explains how to enter data in Full Monte.

Entering data

To enter Full Monte data for individual tasks, you must first open a project and go to the **Add-Ins** ribbon.

> *NOTE: Here and throughout the remainder of this user guide, the instructions are written from the perspective of the Project 2010 user. If you are using Project 2007, there is no concept of ribbons and no icons for Full Monte functions. Instead, all the Full Monte options will appear on the Full Monte drop-down menu.*

Then you can either:

• Select the task or tasks in Project, and then click on the **Full Monte Edit** icon, *or*

• Just click on the icon and select the task(s) that you want to change from within the dialog box.

Whichever option you choose, you will see the Task Information dialog box, shown in figure 3.3. But first, Full Monte does some checks.

Before the edit dialog box opens — and, in fact, before doing other operations on the project — Full Monte makes some checks on the project mappings and may display one of two dialog boxes. If there are no mappings for the active project, which will be the case if this is the first time it has been used in Full Monte, it will check that there is no data in the fields Full Monte is configured to use by default. If there is, it will display the dialog box in figure 3.1.
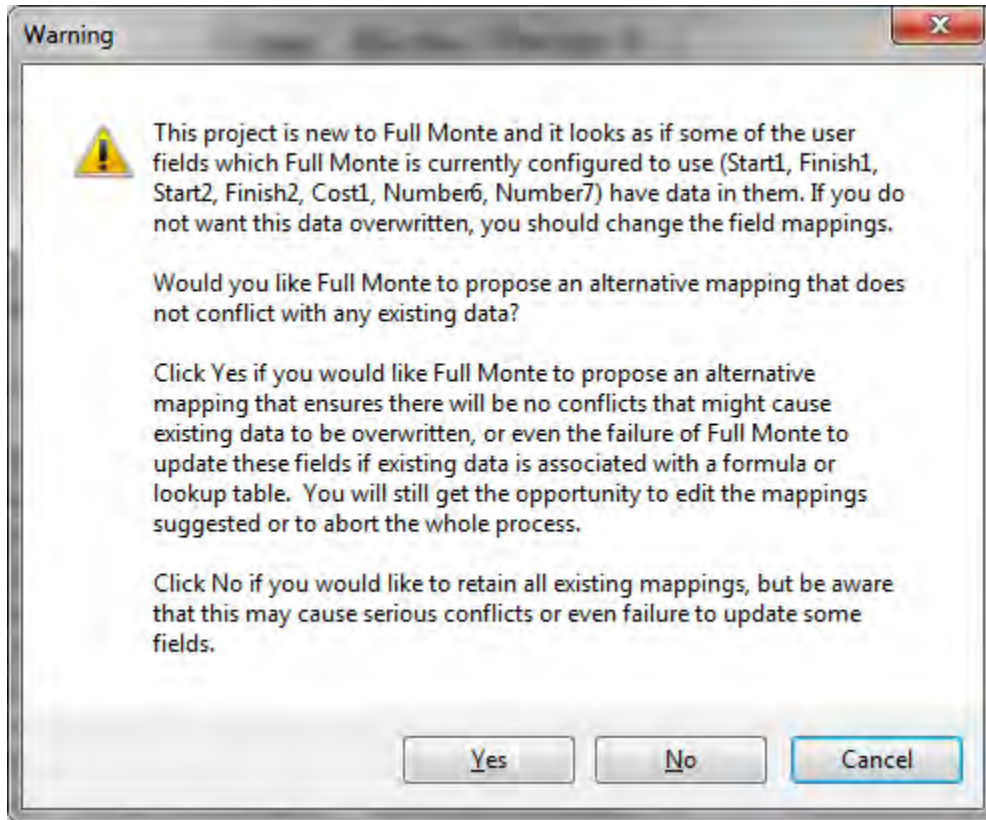
Figure 3.1

The data already in these fields may have been entered directly by you or by another add-in product, and presumably you do not want it overwritten by Full Monte. If you click on **Yes**, the program will take you to the Field Mapping dialog box discussed in Chapter 2, but with nonstandard mappings chosen so there are no collisions with any existing data. **You may alter these mappings. If you do, Full Monte will not check again for data collisions.**

> NOTE: One or more fields may be mapped to **(None)**. This is especially likely for duration fields, since Full Monte requires nine of these and Project provides only 10.

If, on the other hand, the project does have mappings but they do not match the default mappings, you will get the dialog box in figure 3.2.
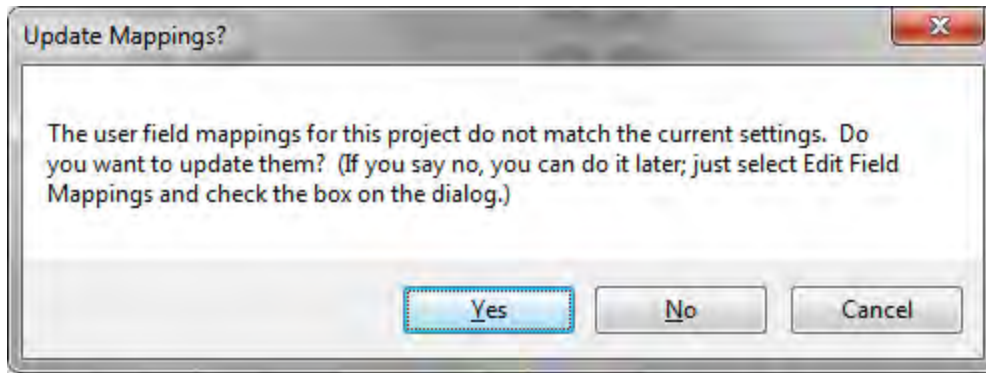
Figure 3.2

This is not really an error, because Full Monte supports different mappings on different projects. If you make all of your mappings the same, it might be less confusing. If you click on **Yes**, Full Monte will not only change the mappings, it will also move the existing data into the new fields. Because this is not an error, the message is not shown when the user accesses a report (because it is read-only), and it is shown only once per session for any particular project.

> *NOTE: Additional dialog boxes may appear in the case of a master project with external subprojects. These are dealt with in Chapter 7, External Subprojects.*

Now look at the edit dialog in figure 3.3.

Figure 3.3

In this case, task 64 was already selected in Project and comes up selected in the dialog box. You can also navigate to different tasks using the left-hand part of the dialog box. Do this by using the up and down arrows on your keyboard or by typing either the task number or task description at the top.

The data in the rest of the dialog box always reflects the task currently selected on the left side. Once you have made any changes to this data, you cannot navigate to a different task until you have either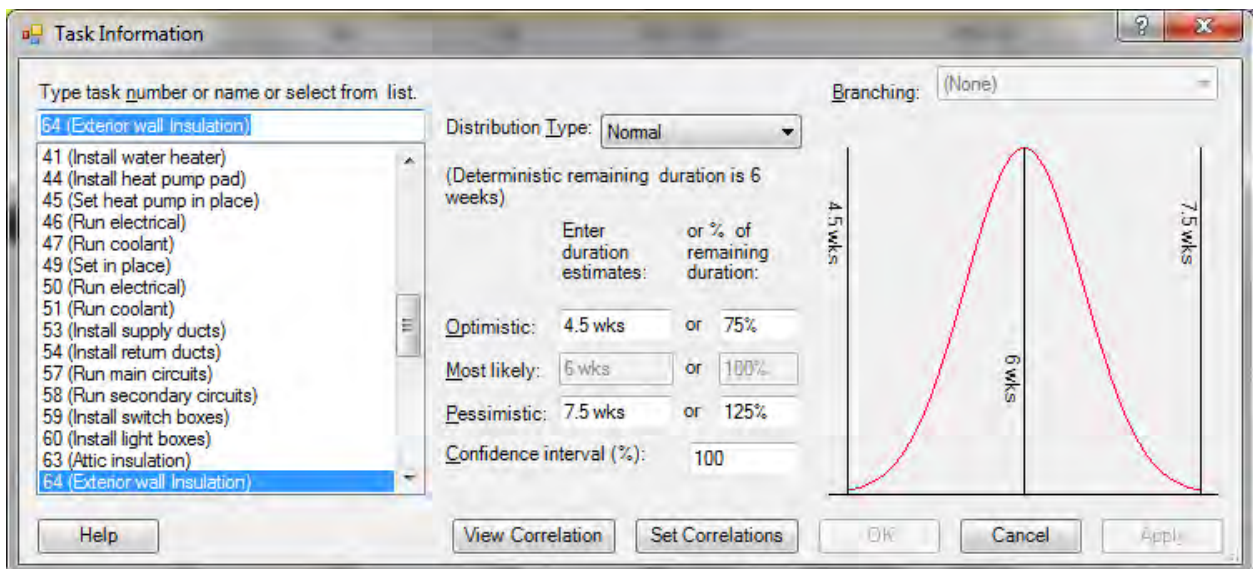 committed those changes by clicking on **Apply** (in the lower-right corner of the screen) or discarded them by pressing **Cancel**. Clicking **OK** also commits the changes but dismisses the dialog box.

To assign a probability distribution to the duration of this task, first choose the general shape from the drop-down menu. Five different shapes are available: beta, lognormal, normal, triangular, and uniform. To determine which to choose for your application, see Appendix A, Probability Distributions, as well as the discussion in Chapter 1, Introduction.

Depending upon which probability distribution shape you choose, the controls will be enabled to allow entry of either a 2- or 3-point estimate of the duration. The uniform, normal, and lognormal distributions are each defined completely by two parameters, so Full Monte calculates the most likely value for you. In either case, you can enter the information either as an *absolute duration* or a *percentage of the deterministic duration*. (You are reminded of the current value of the deterministic duration above the data entry fields.) Whichever entry method you use, Full Monte will prorate the values when the task is progressed or its duration has changed, so you will not have to update these unless you want to.

You can also enter a confidence interval for the optimistic and pessimistic values you enter. By default it is set to 100% , meaning you are 100% sure that the actual value will fall between the values entered.

> *NOTE: In the case of the triangular distribution, this is similar to what other systems call the trigen distribution. In Full Monte, this concept is extended to apply to all distribution types.*
>
> *100% confidence cannot be strictly true for those distribution types that have infinitely long tails, namely the normal or lognormal. In these cases, Full Monte interprets 100% as a 6-standard-deviation range, or about 99.7%.*

As you enter the data described above, the picture will change to indicate what the distribution function actually looks like, with the optimistic, most likely, and pessimistic values marked, along with the absolute limits if these are different.

You can optionally enter correlation information. Full Monte implements correlations differently than other systems, allowing you to correlate each task not directly with other tasks but with outside factors, which are called "correlation sources" in Full Monte. (For more information about correlation sources, please see Appendix B, Correlations.) Click on **Set Correlations** (on the bottom of the screen near the middle) to bring up the correlation grid as shown in figure 3.4.
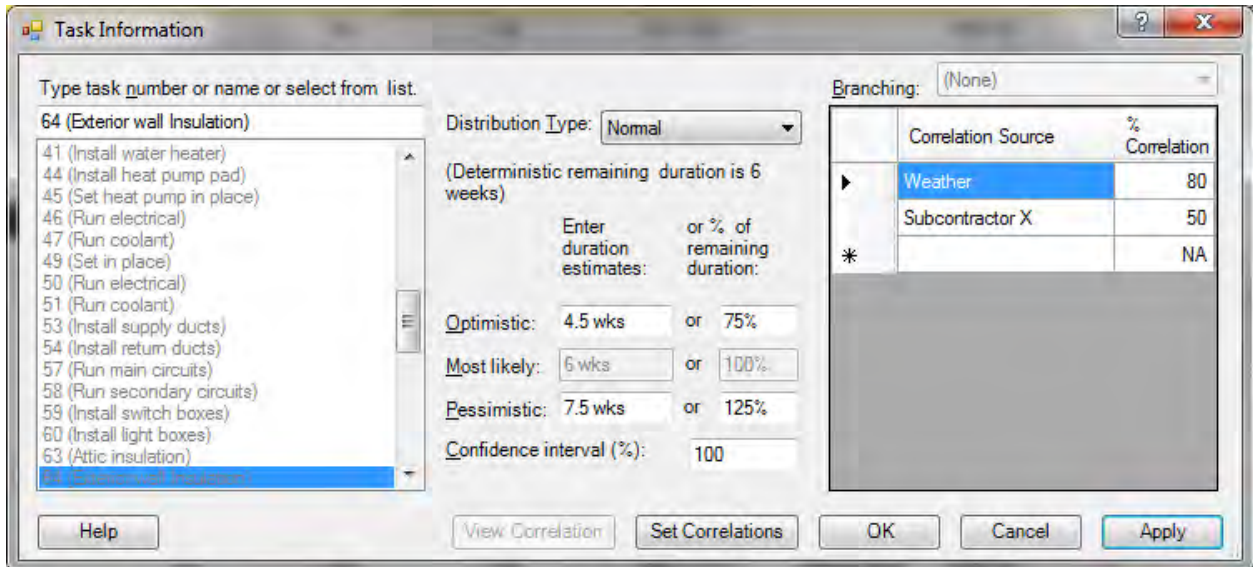
Figure 3.4

Position the cursor in the first cell (on the right side of the screen). From the drop-down menu, choose the correlation source or type in a new one. (Once you type in a correlation source, it will appear in the drop-down menu next time.) In the second column, enter a percentage for the correlation coefficient. This should be between -100% and +100%, and the sum of the squares of the values entered must not exceed 10,000 (i.e., if expressed as fractions rather than percentages, the sum cannot exceed 1).

Note that this correlation coefficient is between the duration of the current task and the specified correlation source. The correlation coefficient between any pair of tasks correlated with this same factor will be the product of the two entered correlation coefficients, interpreting them as fractions. For example, if the two tasks have coefficients of 50% and 70%, the correlation between the two task durations will be 35%.

> *NOTE: This also means that if you enter negative correlations for two tasks, they will be positively correlated with one another. If you want a negative correlation between two tasks, one should be negatively and one positively correlated with the source.*

This novel way of specifying correlations allows Full Monte to properly model correlations between any number of tasks without the user having to resort to programming. (For more information on how this works, see Appendix B, Correlations.) You can see the implied correlations with other tasks after you have applied the changes by clicking **View Correlations**, which brings up the grid illustrated in figure 3.5. (This button is enabled only if there is at least one other task correlated with the current task.)
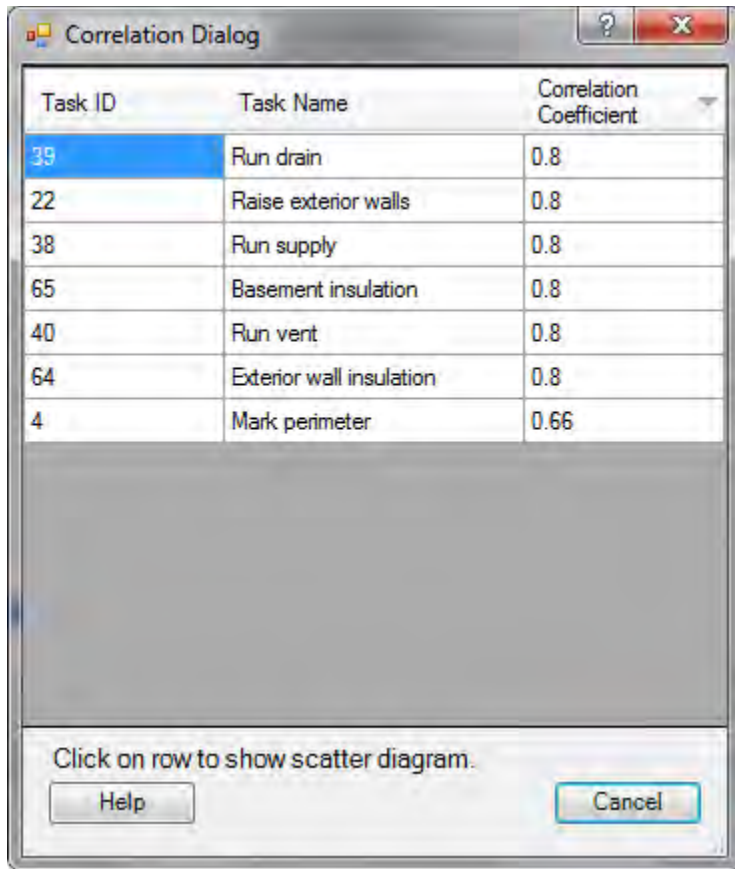
Figure 3.5

The grid shown in figure 3.6 lists all the tasks correlated with the current task, in descending order of the theoretical correlation coefficient between each task and the current task. Clicking on one of the rows produces a scatter diagram of the relationship between the two tasks, as illustrated in figure 3.6.
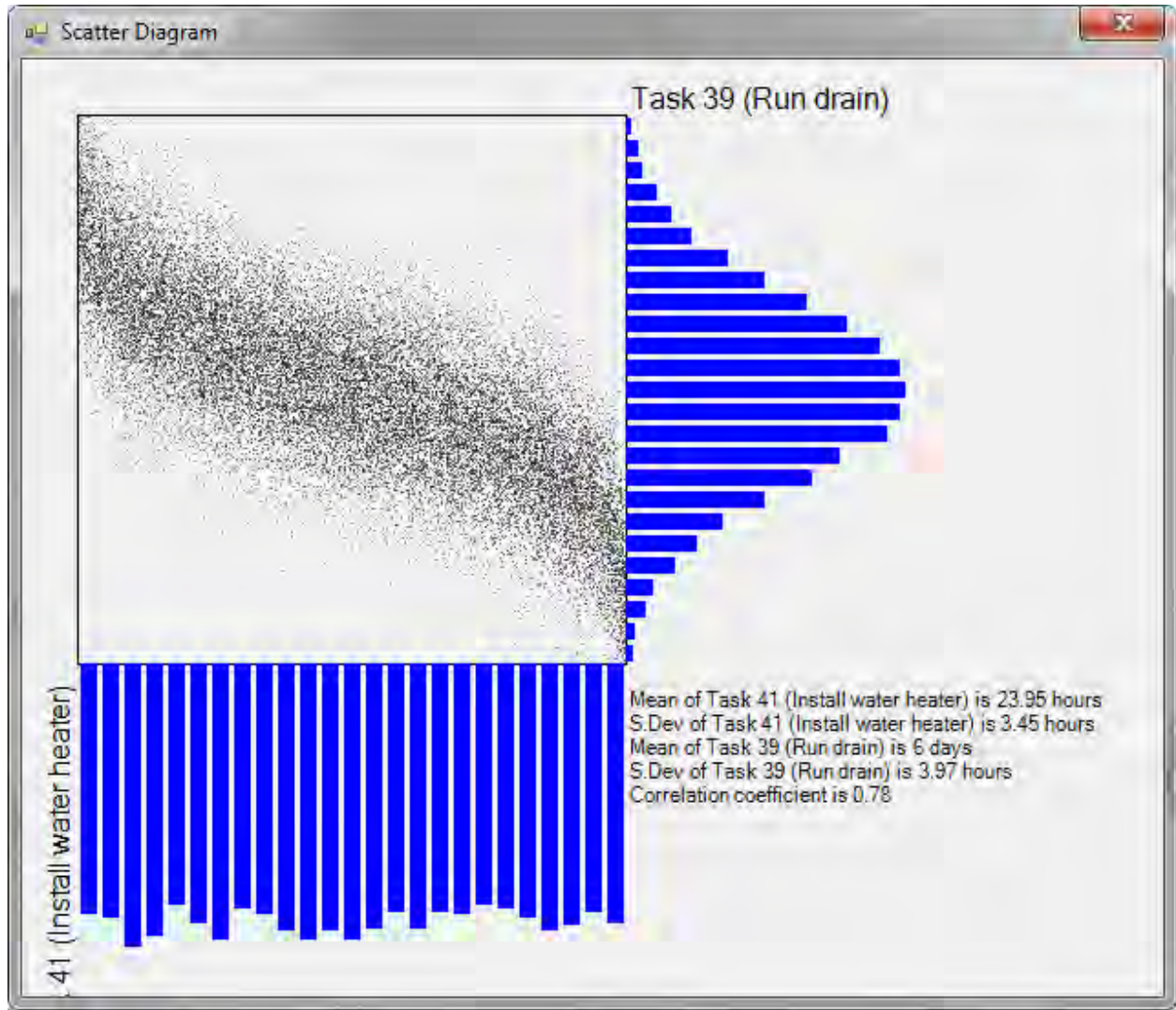
Figure 3.6

Each point on the scatter diagram represents a pair of duration values sampled from the joint distribution of the durations of the two tasks. The blue lines indicate the sampled frequency functions. The text indicates the values of their means and standard deviations and the actually sampled correlation coefficient.

<u>Branching</u>

Branching is a technique whereby the finish-to-start successors to a task are considered mutually exclusive alternatives rather than parallel tasks. Full Monte supports two types of branching, conditional and probabilistic. You can choose which kind to use in the drop-down menu at the top right. (This feature is only enabled if the current task has at least two finish-to-start successors.)

NOTE: The implications of branching are quite complex and the technique should be used sparingly. For details, see Appendix C, Branching.

Depending upon which type of branching you choose, the right side of the dialog box will display one of two different grids. Each grid has a row for every finish-to-start successor.

Conditional branching

For *conditional branching*, the choice of active successor depends on the finish date of the current task. The grid you'll be working in is shown in figure 3.7.
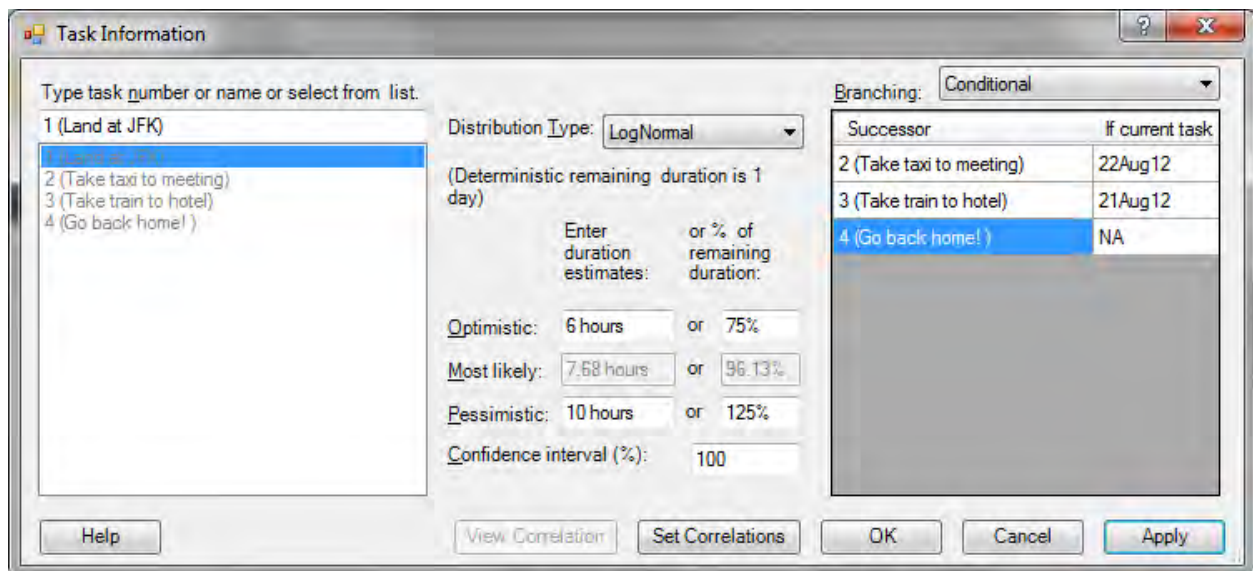
Figure 3.7

Enter dates in all but one of the rows. Every date should be different. Using the above as an example, the data is interpreted as follows: If task 1 (arriving at the airport) finishes on or before August 21, then choose task 2 (take the train to the hotel). Otherwise, if task 1 finishes before August 22, choose task 3 (take a taxi straight to your meeting). Otherwise, choose task 4 (go back home!).

*NOTE: You do not have to enter the dates in order, and the missing date can be anywhere. When you click on **Apply**, Full Monte will display them in order as shown in figure 3.8.*
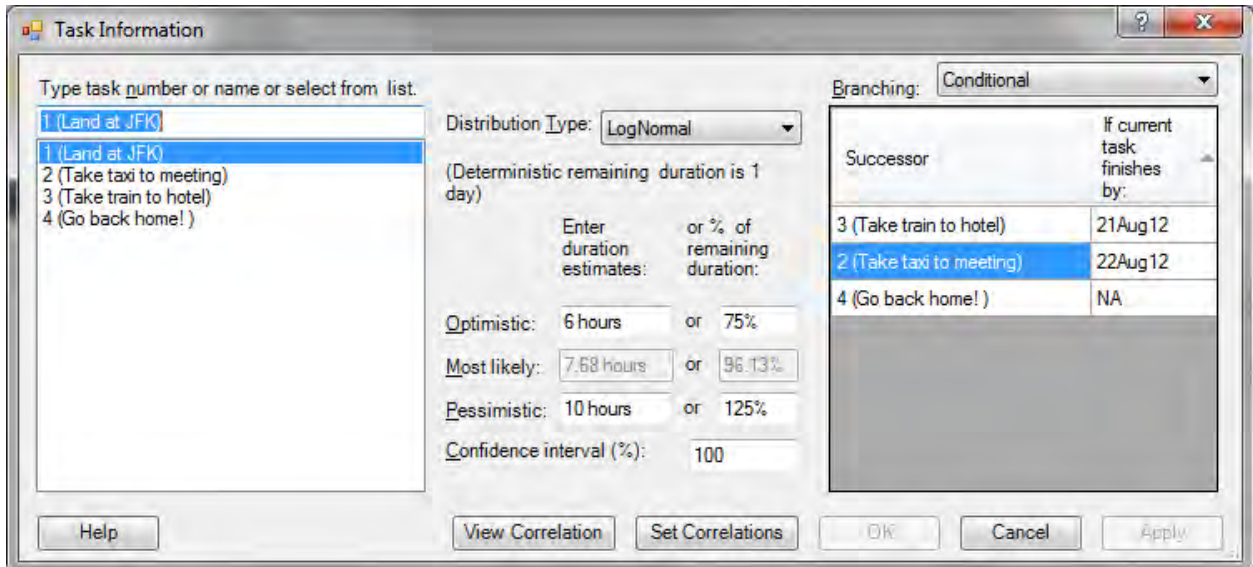
Figure 3.8

Probabilistic branching

In the case of *probabilistic branching*, the choice depends purely on chance. The grid you'll be working in is shown in figure 3.9.
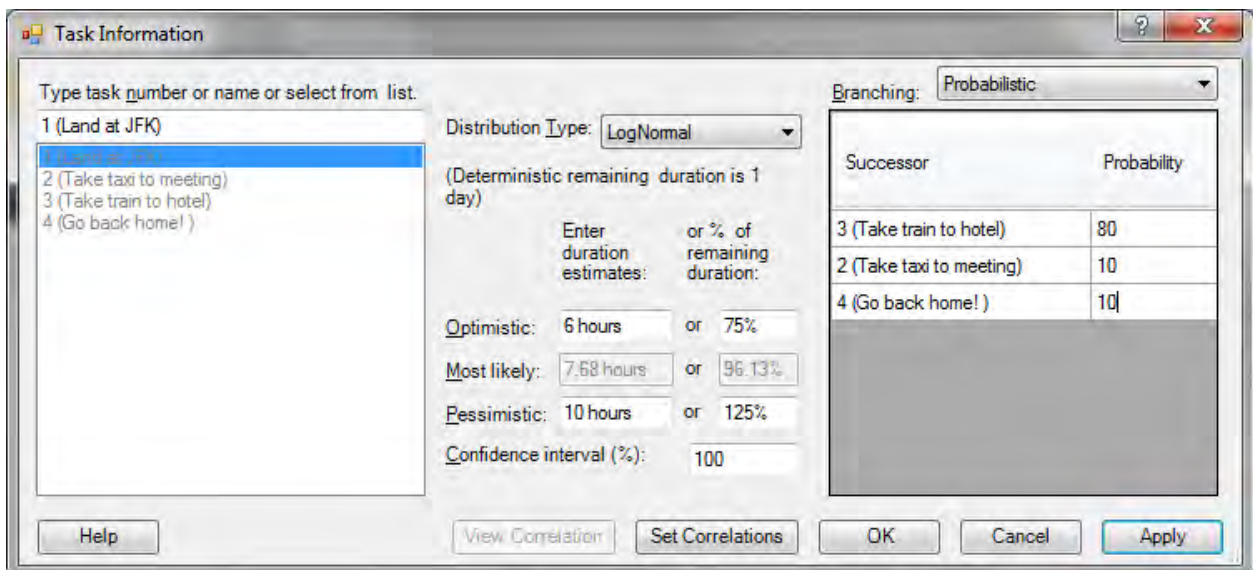
Figure 3.9

Enter relative probability in each row. They should all be greater than zero, but they do not need to add up to anything in particular because they will be scaled. A good practice would be to use percentages and have them add to 100, but this is not required by the software.

When you have finished entering data for the current task, you can click on either **OK** or **Apply**.

- **OK** will update the data and close the dialog box.

- **Apply** will update the data but leave the dialog box open so you can update further tasks. You cannot navigate to another task while there are pending changes to the current one.

The **Cancel** button has a different effect depending on whether there are pending changes to the current task, which you can easily tell by whether the **OK** and **Apply** buttons are enabled. If these buttons are enabled, there are pending changes and pressing the **Cancel** button will discard the changes, reset the data to its original values, and disable the **OK** and **Apply** buttons. Otherwise the **Cancel** button closes the dialog box.

Global edit dialog box

An alternative method of data entry is the Global Edit dialog box, accessed from the Global Edit icon on the add-in ribbon. To apply global edits, you must first open a project. It is a good idea to save the current state of the project before doing this in case you make a mistake.

We do not recommended using global edits for all tasks. Tasks that represent significant risk should be assessed individually. However, global edits can be a useful tool to:

- Fill in a background low level of risk for the remaining tasks, *or*

- Enter values for a set of tasks that are similar in their risk characteristics. In this case, select the tasks you want to change in Project before invoking the Full Monte Global Edit dialog box, as seen in figure 3.10.
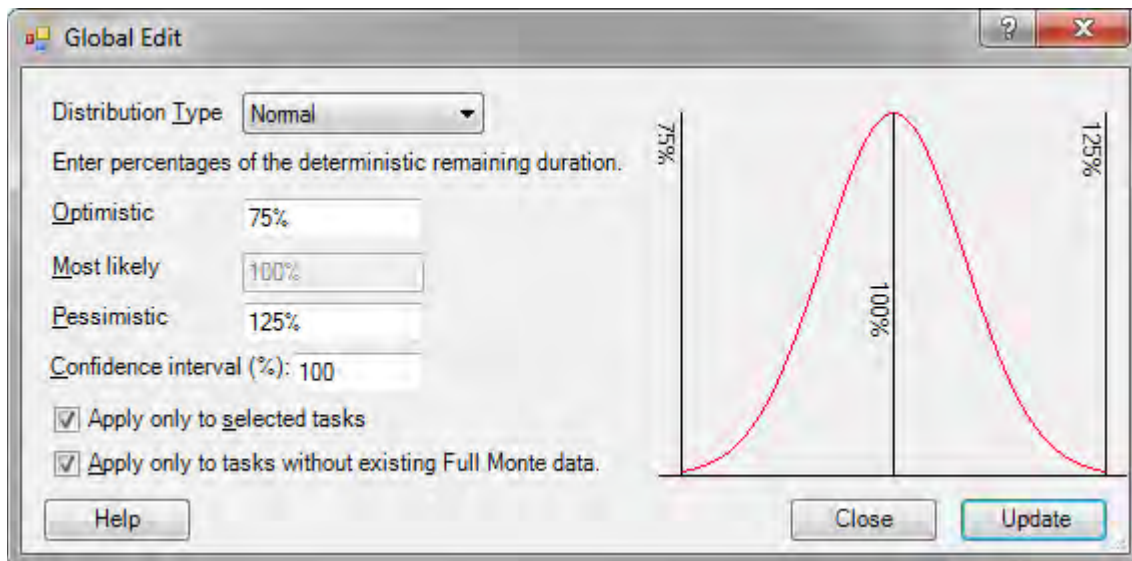
Figure 3.10

NOTE: The Apply only to selected tasks *checkbox will be disabled if there are less than two tasks selected.*

Choose a distribution type from the drop-down menu, and then enter the 2- or 3-point estimates and the confidence interval. Since the durations of the task to be changed will in general be different, these must be entered in percentages.

NOTE: You cannot globally update correlation or branching information.

Then make sure you have the checkboxes set as you want them. Click on **OK** to do the global update. Full Monte will tell you how many tasks it has changed, as shown in figure 3.11.
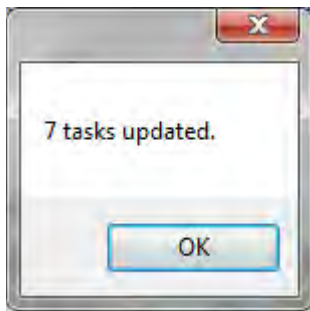
Figure 3.11

If this is not what you were expecting, you might consider not saving the project and trying again.

CHAPTER 4: Analyzing the Project

Once you have entered the data, you will be ready to analyze and interpret it. This chapter shows you how with various types of analyses and reports available in Full Monte.

Risk and sensitivity analysis

Full Monte performs both risk and sensitivity analyses in one operation. After you click on **Analyze**, you'll see the Risk Analysis dialog box, shown in figure 4.1.

NOTE: Full Monte first checks the field mappings as described in Chapter 2, Getting Started.
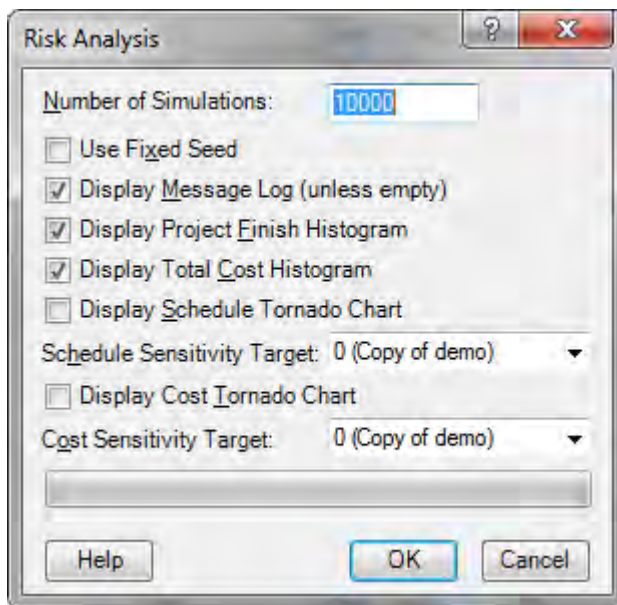
Figure 4.1

From this screen, you can choose a number of options, as described below.

Number of Simulations: The larger this number, the better your results will be. Not only will your results be more accurate, but the graphics will tend to be smoother and therefore more aesthetically pleasing. The only downside to doing more simulations is the additional processing time. However, Full Monte is fast, so you can afford to do more simulations than might be practical with slower products.

Use Fixed Seed: Risk analysis uses what are called pseudorandom numbers. They are not really random, because if you know the algorithm you can predict the next value with perfect precision. But they appear random and pass statistical tests for randomness. The random number generator starts with what is called a seed value, which by default is determined by the date and time. This means that if you do the same risk analysis twice, you will get the same results only to the statistical accuracy of the

simulation. Unless you do a very large number of simulations, you will not always get *exactly* the same results. If you want exact repeatability, check this box and the random number generator will be initiated with the same seed value every time.

Display Message Log: If an anomalous data is found during the process, a message is output to the message log field. The message log can be viewed at any time, but if you want to see it when the analysis is complete, check this box. The report will not be shown if there are no messages, and the checkbox will be disabled if the **analysis log messages** field is not mapped to a custom text field.

Display Project Finish Histogram: Full Monte always calculates histogram data for the early finish of the whole project, and this can be displayed at any time. This option displays it automatically when the analysis is complete. If you are using correlations, there is additional functionality available when displayed as part of the analysis process, as described in Chapter 5, Investigating Correlation Sensitivity.

Display Total Cost Histogram: Full Monte always calculates histogram data for the total cost of the project, and this can be displayed at any time. This option displays it automatically when the analysis is complete. This option is disabled if there is no cost data. If you are using correlations, there is additional functionality available when displayed as part of the analysis process. This is explained further in Chapter 5, Investigating Correlation Sensitivity.

Display Schedule Tornado Chart: This bar chart shows those tasks to which the project's early finish date (or other sensitivity target; see next paragraph) is most sensitive. The bars show an estimate of the difference between the expected project early finish if that task realizes its shortest possible duration and if it realizes its longest possible duration. By default, the report is sorted in descending order of the length of the bars. It displays only those tasks that have a statistically significant effect on the project finish. Two other measures of sensitivity, Sensitivity Index and Sensitivity Attributable Variance, are also displayed in the tabular part of the report. You can display the bar chart at any time, but there is additional functionality when displayed as part of the analysis process, as noted in Chapter 6, Investigating Task Sensitivity.

Schedule Sensitivity Target: By default, this is the project finish. It is also possible to measure sensitivity relative to any milestone. If you select a milestone, all schedule sensitivity results will relate to the sensitivity of the early date for this milestone to the task in question. For more on sensitivity, see Chapter 6, Investigating Task Sensitivity.

Display Cost Tornado Chart: This bar chart shows those tasks to which the project's cost (or other sensitivity target; see next paragraph) is most sensitive. The bars show an estimate of the difference between the expected cost if that task realizes its lowest possible cost and if it realizes its highest possible cost. By default, the report is sorted in descending order of the length of the bars. It displays only those tasks that have a statistically significant effect on the project cost. You can display the bar chart at any time, but there is additional functionality when displayed as part of the analysis process, as noted in Chapter 6, Investigating Task Sensitivity.

Cost Sensitivity Target: By default, this is the total project cost. It is also possible to measure sensitivity relative to any summary task. If you select a summary task, all cost sensitivity results will relate to the sensitivity of the cost of this summary task to the task in question. For more on sensitivity, see Chapter 6, Investigating Task Sensitivity.

When you have selected the options you want:

- Press **OK** to start the analysis. The progress bar roughly indicates the proportion of the analysis that is complete.

- Pressing **Cancel** before pressing OK will dismiss the dialog box.

  *NOTE: Pressing **Cancel** after pressing OK but before the analysis is complete will stop the analysis, though on a large project you may have to wait a few seconds for this to take effect.*

When the analysis is complete, it will display the reports you requested. As noted above, these offer additional functionality described in the next two chapters, Investigating Correlation Sensitivity and Investigating Task Sensitivity.

CHAPTER 5: Investigating Correlation Sensitivity

This chapter deals specifically with correlation sensitivity accessed from this report. Other features of the histogram report are described in Chapter 8, Customizing Reports.

Investigating the effect of correlation sources

If you have defined correlation sources, then it is possible to investigate the effect of each of these sources, separately or in various combinations, from the histograms produced when you check the appropriate box on the analysis dialog box described in the previous chapter. These histograms look like the example shown in figure 5.1.
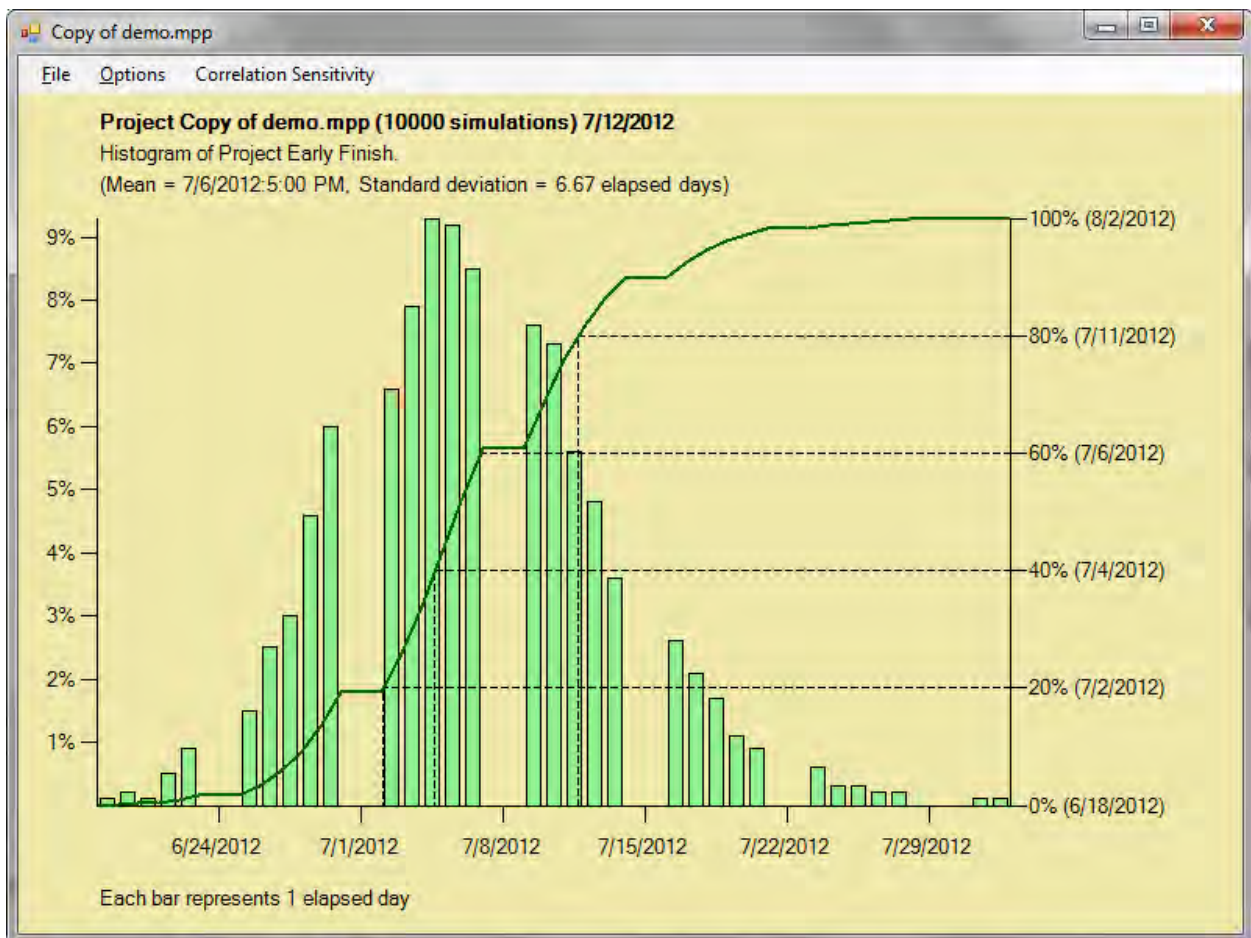
Figure 5.1

Clicking on **Correlation Sensitivity** (at the top of the screen) brings up the dialog box in figure 5.2.
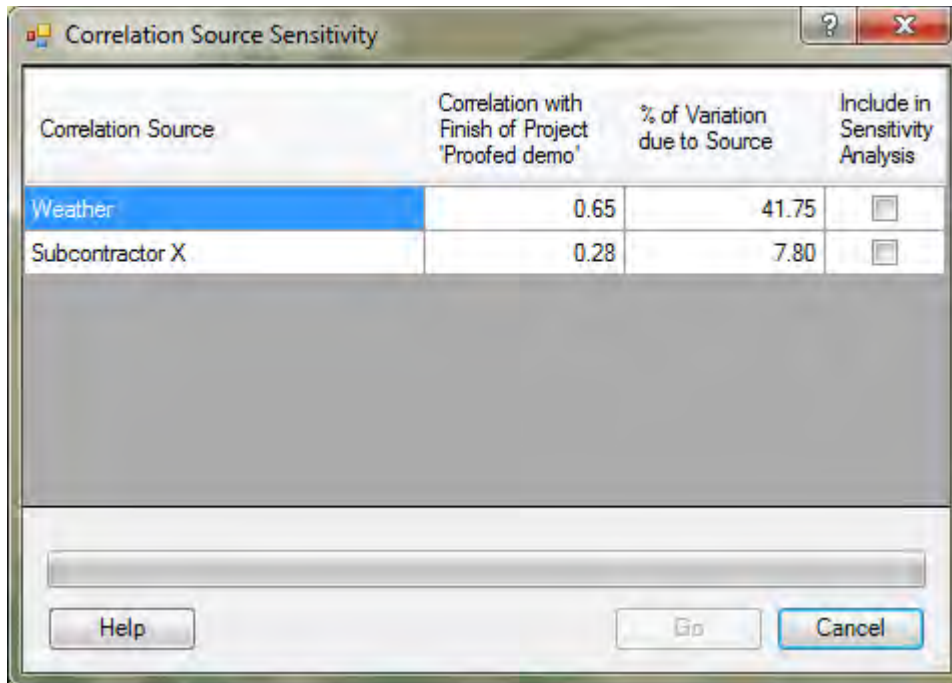
Figure 5.2

The Correlation Source Sensitivity dialog box displays estimates of the sensitivity of the project finish date (or other sensitivity target) to each of the correlation sources. Checking one or more of the checkboxes enables the **Re-run** button. Clicking on this button will re-run the simulation twice, with the checked correlation source(s) set to their minimum and maximum values, respectively. The results are displayed as two cumulative distribution functions as illustrated in figure 5.3.
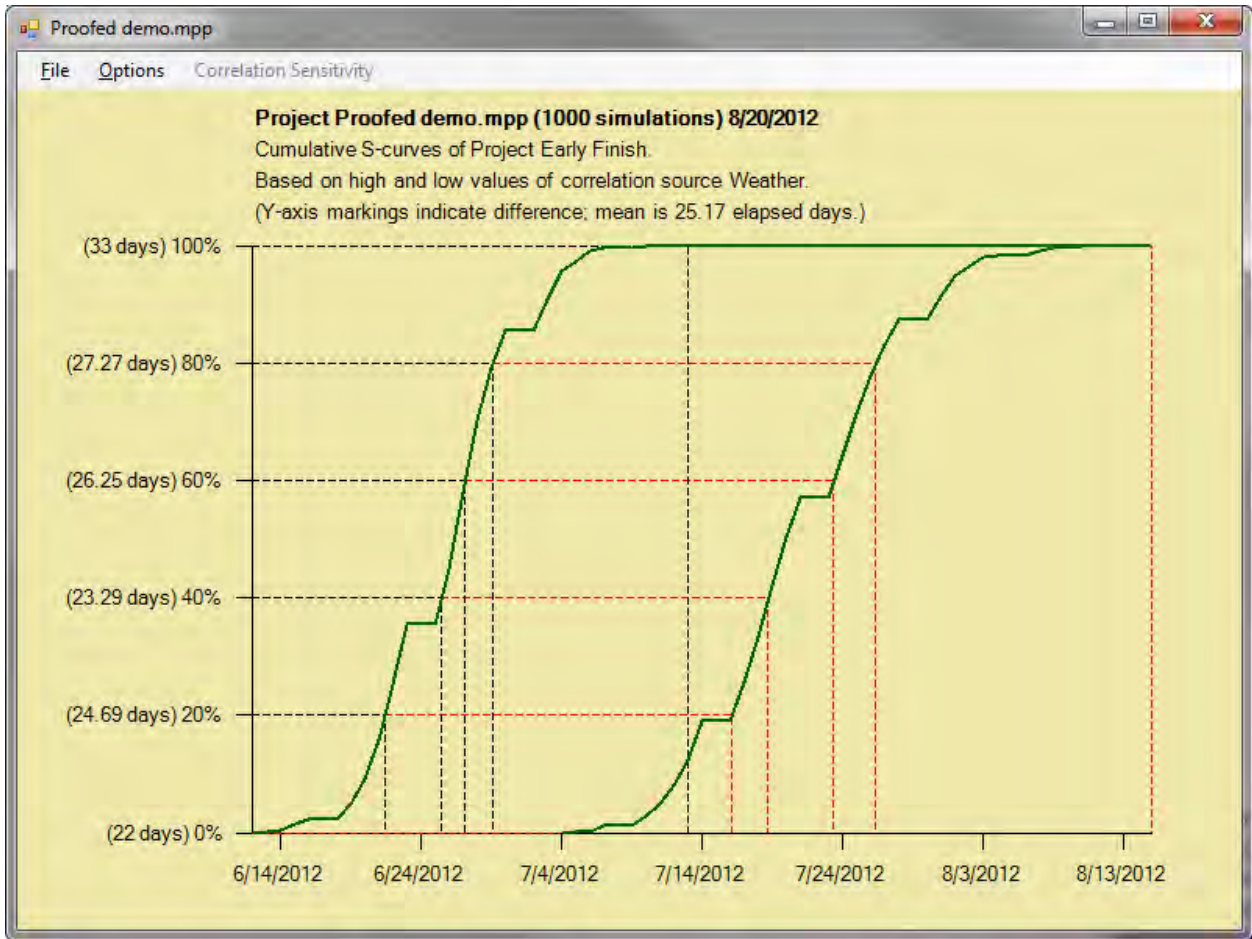
Figure 5.3

> *NOTE: This report will not be saved. It can, however, be printed or copied and pasted into another application or reproduced by re-running the analysis.*

The double S-curve report can be dismissed by clicking anywhere on it, which returns you to the dialog box shown in figure 5.2. You can then investigate other correlation sources or combinations of correlation sources. When you have finished, click on the **Cancel** button.

If you requested the cost histogram, this will be displayed next. The correlation sensitivity for cost works the same way as the finish date sensitivity.

If you requested one or both of the tornado charts, these will be displayed next. This is the subject of the next chapter, Investigating Task Sensitivity.

CHAPTER 6: Investigating Task Sensitivity

This chapter continues the process begun in the prior chapter and explains how to investigate task sensitivity.

Using the tornado chart

If you requested a tornado chart for the project finish date (or other sensitivity target), this will be displayed next and will look as shown in figure 6.1. During the simulation, Full Monte makes an estimate of the effect of each task's duration on the project finish. In some cases, for tasks that are never critical, this effect will be zero. But for some tasks it will be significant. For each of these tasks, Full Monte produces a bar stretching between two estimates of the expected early finish of the sensitivity target: one based on the shortest duration of the task in question and one based on the longest duration of this task. All other tasks vary as specified. The report is sorted in decreasing order by the length of the bars, and the dates are displayed in the columns of the report.

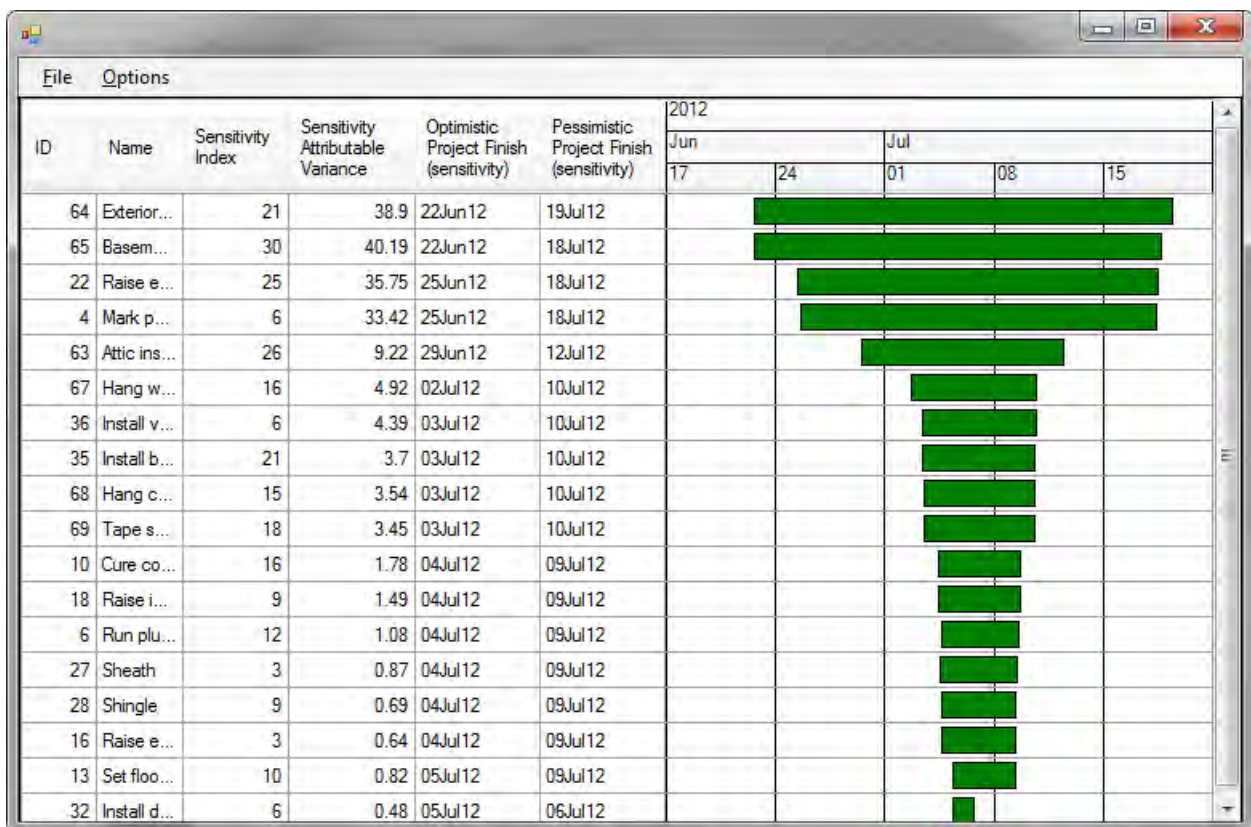| ID	Name	Sensitivity Index	Sensitivity Attributable Variance	Optimistic Project Finish (sensitivity)	Pessimistic Project Finish (sensitivity)	
64	Exterior...	21	38.9	22Jun12	19Jul12	
65	Basem...	30	40.19	22Jun12	18Jul12	
22	Raise e...	25	35.75	25Jun12	18Jul12	
4	Mark p...	6	33.42	25Jun12	18Jul12	
63	Attic ins...	26	9.22	29Jun12	12Jul12	
67	Hang w...	16	4.92	02Jul12	10Jul12	
36	Install v...	6	4.39	03Jul12	10Jul12	
35	Install b...	21	3.7	03Jul12	10Jul12	
68	Hang c...	15	3.54	03Jul12	10Jul12	
69	Tape s...	18	3.45	03Jul12	10Jul12	
10	Cure co...	16	1.78	04Jul12	09Jul12	
18	Raise i...	9	1.49	04Jul12	09Jul12	
6	Run plu...	12	1.08	04Jul12	09Jul12	
27	Sheath	3	0.87	04Jul12	09Jul12	
28	Shingle	9	0.69	04Jul12	09Jul12	
16	Raise e...	3	0.64	04Jul12	09Jul12	
13	Set floo...	10	0.82	05Jul12	09Jul12	
32	Install d...	6	0.48	05Jul12	06Jul12	

Figure 6.1

NOTE: These dates are estimates. The tornado chart offers a rough guide to the tasks that may need further investigation. The only way to reliably establish sensitivity is to run two full simulations (explained below).

Because the bars represent estimates, Full Monte also displays two additional guides in this report:

Sensitivity index. The sensitivity index is defined as the percentage of time the task is critical, multiplied by the standard deviation of its duration and divided by the standard deviation of the finish date of the sensitivity target. This can be a useful alternative measure since the rank order differs based on the length of the bars. However, if the task durations are correlated, the report can give spurious results.

Sensitivity attributable variance. The percentage of variance attributable to the task is the square of the correlation between the task duration and the finish date of the sensitivity target, expressed as a percentage. In simple cases, these figures will add up to 100%, but not when there are correlation sources involved. The rank order of this measure generally compares closely with that of the length of the bars.

As with the correlation sensitivity discussed in the preceding chapter, these measures are rough guides to indicate which tasks need more thorough investigation.

To more thoroughly investigate the sensitivity to the duration of a particular task, click on the corresponding bar in the tornado chart. The dialog box in figure 6.2 will appear. Click on **Go** to run two full simulations, one with the task duration set to its optimistic value and one with it set to its pessimistic value. All other task durations vary as specified.
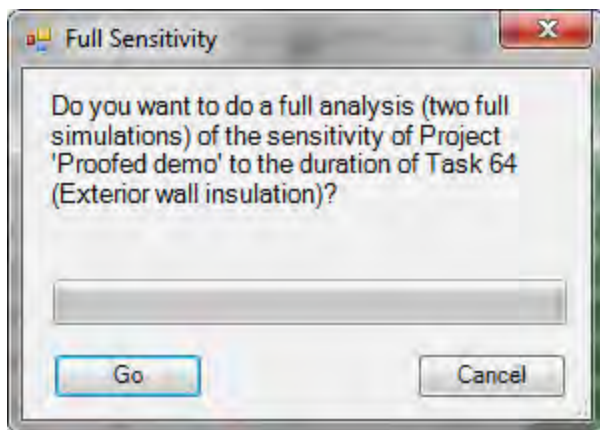

Figure 6.2

Full Monte will display the results as two cumulative distribution functions, or S-curves, as illustrated in figure 6.3.
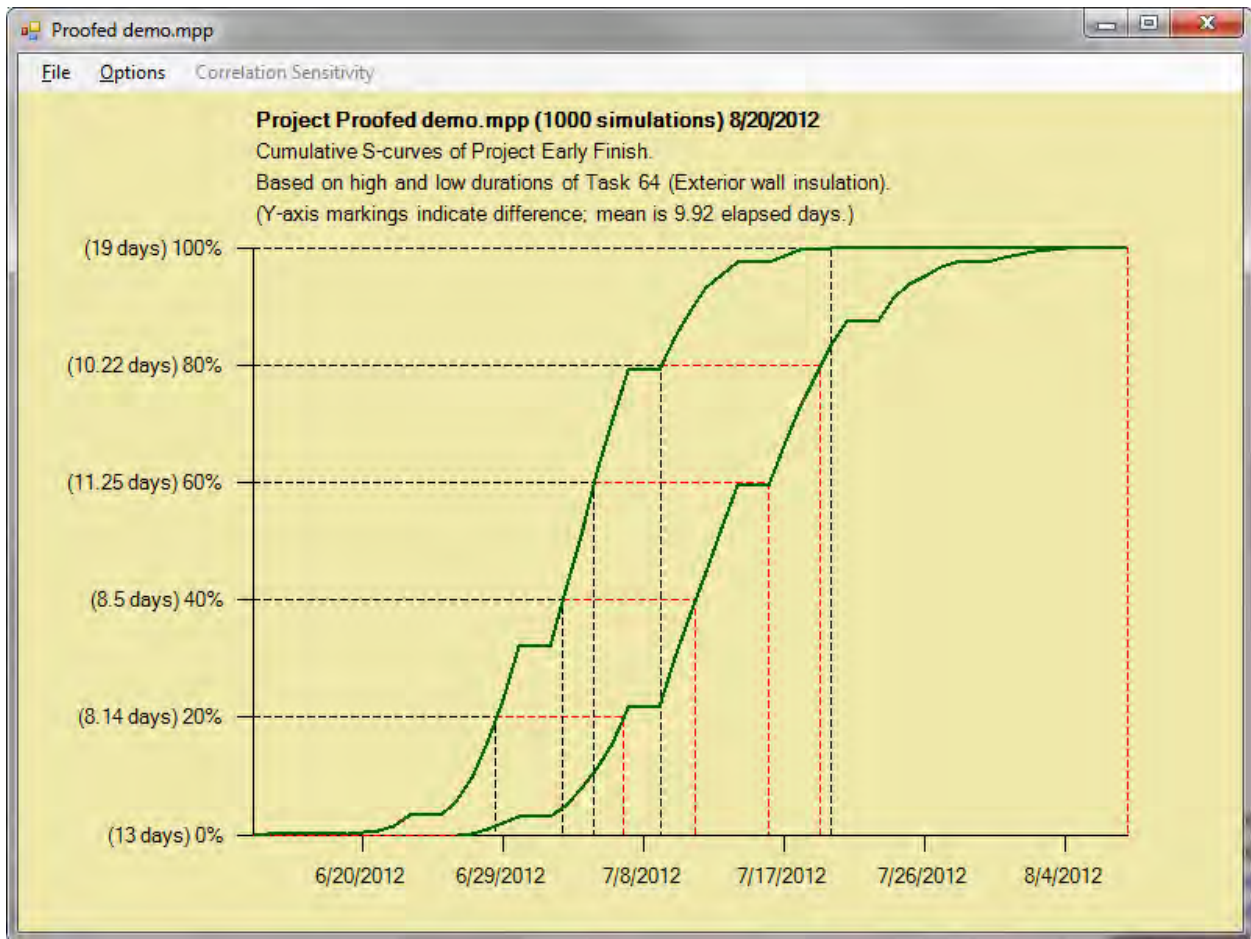
Figure 6.3

NOTE: This report will not be saved. It can, however, be printed or copied and pasted into another application or reproduced by re-running the analysis.

The analysis in figure 6.3 is more useful than the tornado chart for two reasons:

• This chart allows you to see the sensitivity of the percentile you are interested in, which can be quite different from the sensitivity of the expected value as estimated in the tornado chart.

• This analysis is more accurate. In this case, the tornado chart has seriously overestimated the sensitivity to individual tasks due to correlations. (The fact that the attributable variances added up to more than 100% should have been a clue that this was the case.) The length of the bar for task 64 is nearly four weeks, whereas the actual sensitivity is more like one to two weeks, depending upon the percentile of interest.

Dismiss the double S-curve report by clicking anywhere on it. You'll then be returned to the tornado chart to investigate other tasks. When you have finished, click on the **red X** button at the top right of the form.

Cost Sensitivity

If you requested the cost tornado chart, Full Monte will display this next. It works the same way as the finish date with a few special considerations:

- Costs are additive whereas task durations are not. This means that the effect on the total cost should be easier to determine. In a simple case, a task that can cost between $750 and $900 should have precisely this effect on the total cost, and it should not be necessary to do a simulation to find this out. In this case, the tornado chart bar should be exactly $150 long. Nevertheless, Full Monte does use the simulation to determine the values for the tornado chart bars. This means there will be some random variation in these values. In the simple case described above, they will not exactly match those values expected. However, the simulation is more reliable in more complex cases where there is escalation or branching.

- To the extent that cost sensitivity is due to escalation, it results primarily from variability in the start date of the task, so it is a result of variability in the durations of predecessor tasks rather than the task in question.

If you did not request the cost tornado chart, you will be returned to the current Project view.

CHAPTER 7: Standard Reports

Full Monte has a range of options for viewing and printing standard reports. To learn how to customize reports, see Chapter 8, Customizing Reports.

Viewing and printing reports

Full Monte calculates and stores the results of your data analyses in Project custom fields. It's possible to see most of your data in Project views like the Gantt Chart view. Exceptions include histograms and S-curves, the data for which are stored in text fields but need interpretation. Full Monte makes this easy by renaming the custom fields to indicate their content.

To view a report, you must first open the project you're working on. Then click on the **Reports** icon. The first time you do this, it will bring up a standard report called Default. (More generally, once you have viewed a report, it will bring up the most recently viewed report.) Default is a simple tabular report, containing a lot of the information calculated by risk analysis.

>   *NOTE: Full Monte first checks the field mappings as described in Chapter 2, Getting Started.*

To change to a different report, click on the **File** menu, then **Open Report Specification**. You will then be presented with a list of the available reports, as shown in figure 7.1.
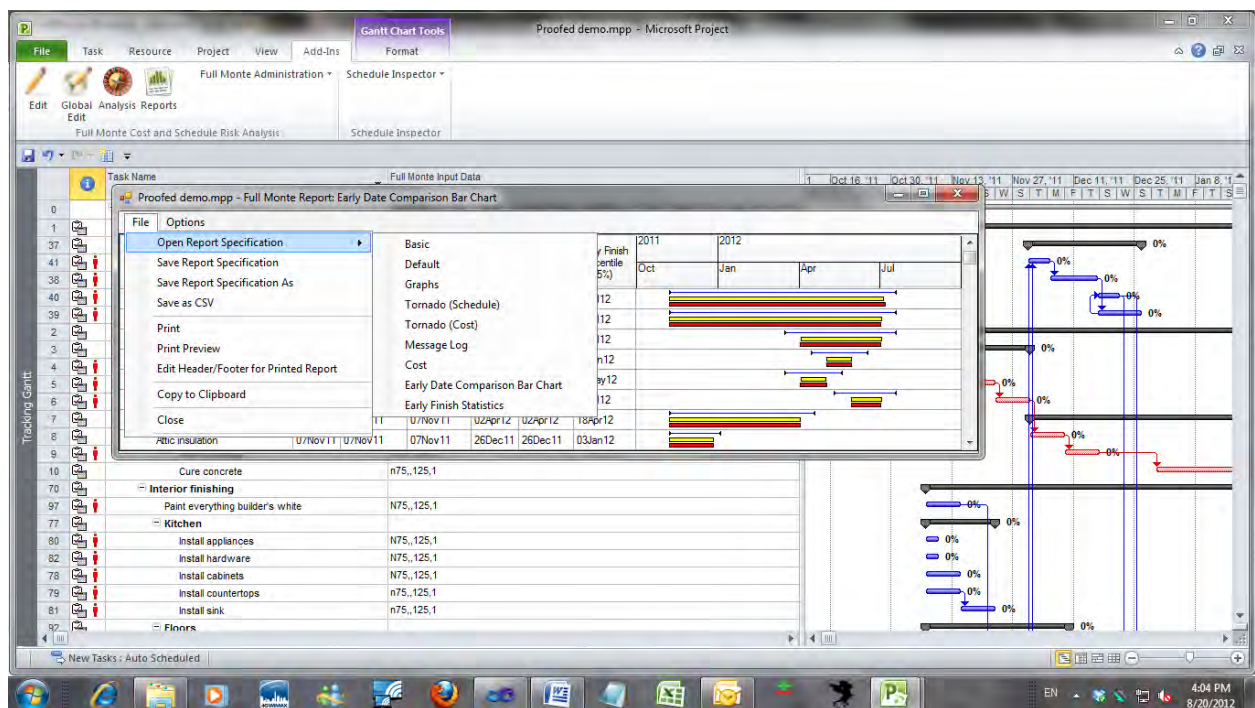

Figure 7.1

Select the report you want to view and click on it.

> *NOTE: All of the reports listed in figure 7.1 are standard reports in Full Monte. The first six are different in that they are hard-coded so you can never lose them. You can modify these reports, but when you save them you will be forced to give them a different name. (In fact, the Basic report has very little information and is intended as a template for you to begin creating your own report.) The rest are similar to reports that you can create yourself.*

To print this or any other tabular report, open it. On the menu, click on **File**, then **Print** (or **Print Preview**, if you want to view the report before printing). Printed reports have optional and customizable headers and footers as discussed in Chapter 8, Customizing Reports.

Double clicking on a data row in a tabular report brings up the Results dialog box for that task, which is shown in figure 7.2. This shows all the values calculated by risk analysis and sensitivity analysis, except for graphical information that is accessed indirectly through buttons on the dialog box.

For an explanation of what the displayed values mean, see Appendix E, Interpreting Full Monte Output Fields.
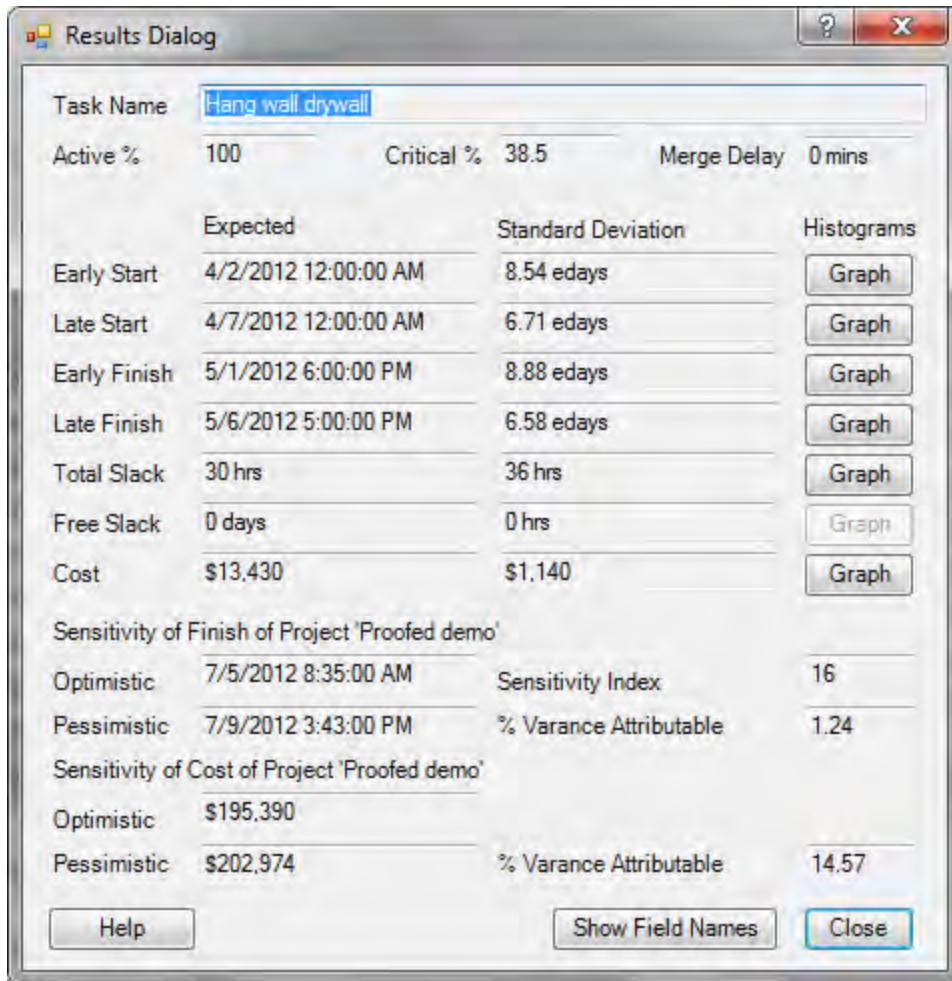
Figure 7.2

The dialog box also has a button that will toggle between showing the results for the task and showing the field mappings for all values calculated by Full Monte, as shown in figure 7.3.
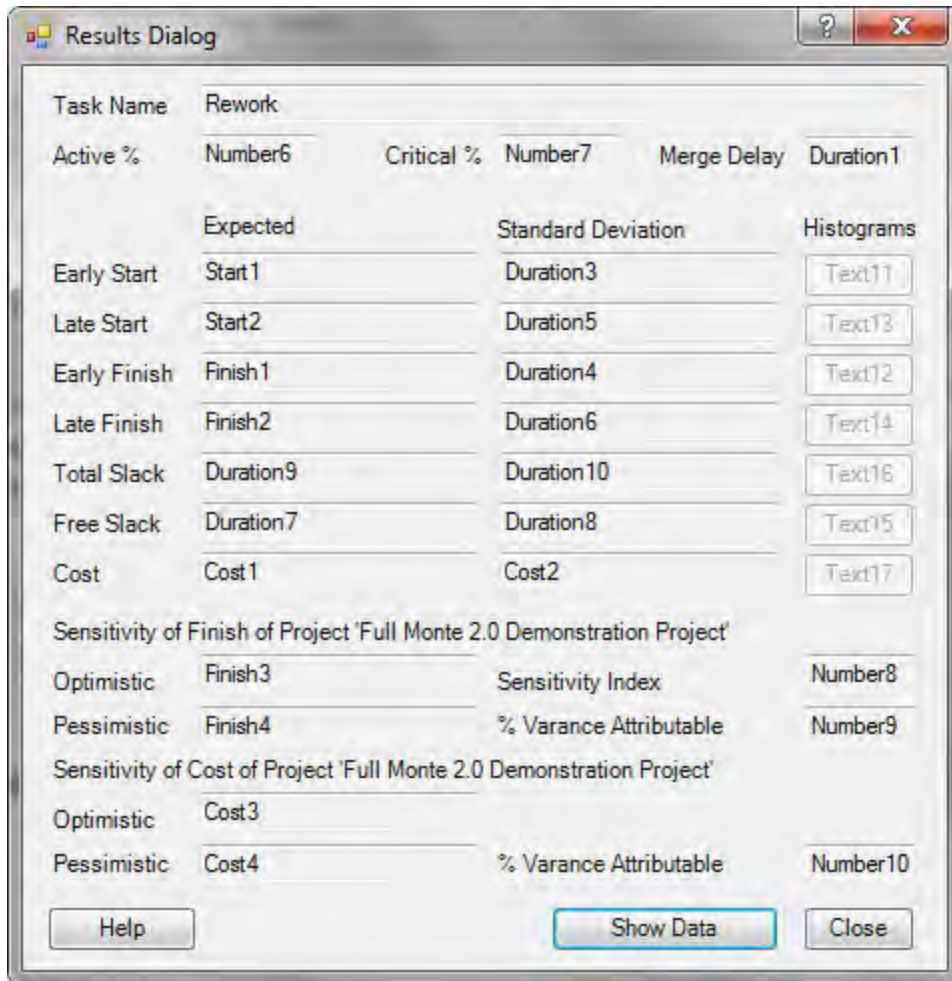
Figure 7.3

Displaying graphs

One of the standard reports is called Graphs, and it is comprised almost entirely of buttons, as shown in figure 7.4. Each of the buttons represents distribution data available to be displayed graphically, as shown in figure 7.5.

To print a graphic report, on the menu strip click on **File**, then **Print** (or **Print Preview**, if you want to view the report before printing).

Click anywhere on the report to close the window.

Figure 7.4

Click on a button to bring up a particular item of data for a particular task, for example, cost or free slack. The window that is displayed will look like that in figure 7.5.
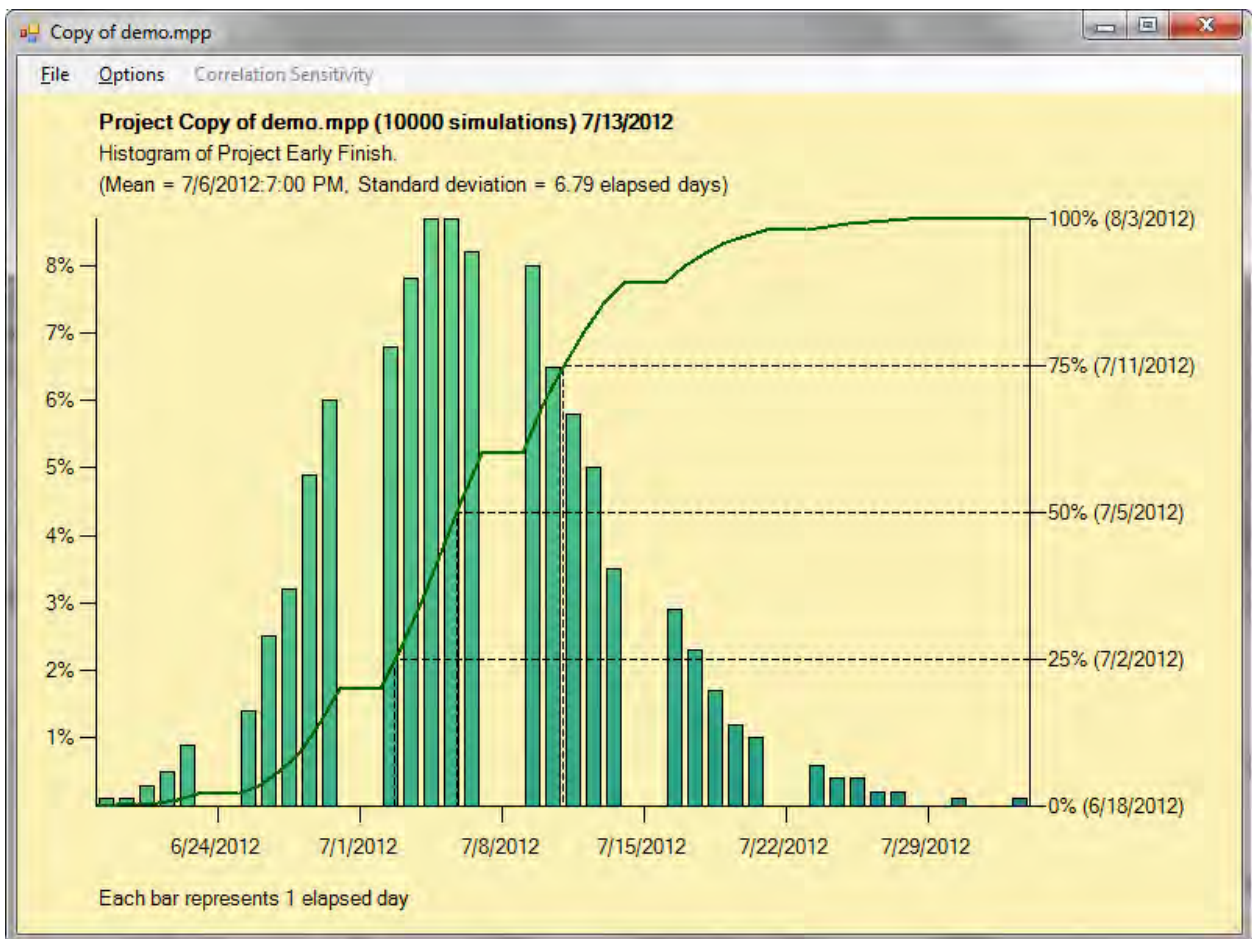
Figure 7.5

The histogram bars represent the probability of the actual value falling in a particular interval, in the chart above, a day. (The gaps are weekends.) The height of the bar is measured against the left-hand axis.

53

The S-curve represents the cumulative probability distribution. In other words, the probability that the actual value will fall to the left of a particular point on the graph. This value always goes from 0 to 100%, and in the chart is measured against the right axis.

NOTE: One possible customization is to exclude the histogram. In this case, the S-curve values would be measured against the left axis.

Values can also be displayed by hovering your cursor over the S-curve or over a histogram bar.

The **<u>File</u>** menu prints this report or copies it to the clipboard. The **<u>Options</u>** menu allows you to customize the report in a number of ways, as described in Chapter 8, Customizing Reports.

CHAPTER 8: Customizing Reports

In addition to the standard reports, users can add their own customized reports.

Tabular reports

Tabular reports can be customized in a number of ways. You can:

- Delete and add columns

- Define bar charts

- Change the order and width of the columns

- Change the sort order

- Freeze columns on the left so they don't scroll

- Change the date format

- Determine whether the project summary is shown

- Determine whether empty rows are shown

- Determine whether the project hierarchy is shown (by indenting the Name column and also by keeping its integrity when data is sorted)

- Change the content of optional header and footer sections when the report is printed

When you are done, you can save the report with a unique name so you can use it later with the same or other projects.

To delete a column: Right-click on the column header you wish to delete. Pick **Delete Column** from the menu. (If you delete all but one column, Full Monte will not allow you to delete the remaining one.)

To add a column: Right-click on the header of the column just before where you want the new column to appear. Pick **Add Column** from the menu. You'll see three choices:

- Full Monte column

- Bar chart column

- Microsoft Project column

Bar charts are dealt with later in this chapter. Selecting either of the other two options invokes a list of supported fields, excluding those that are already in the report, in a series of cascading menus as shown in figure 8.1:
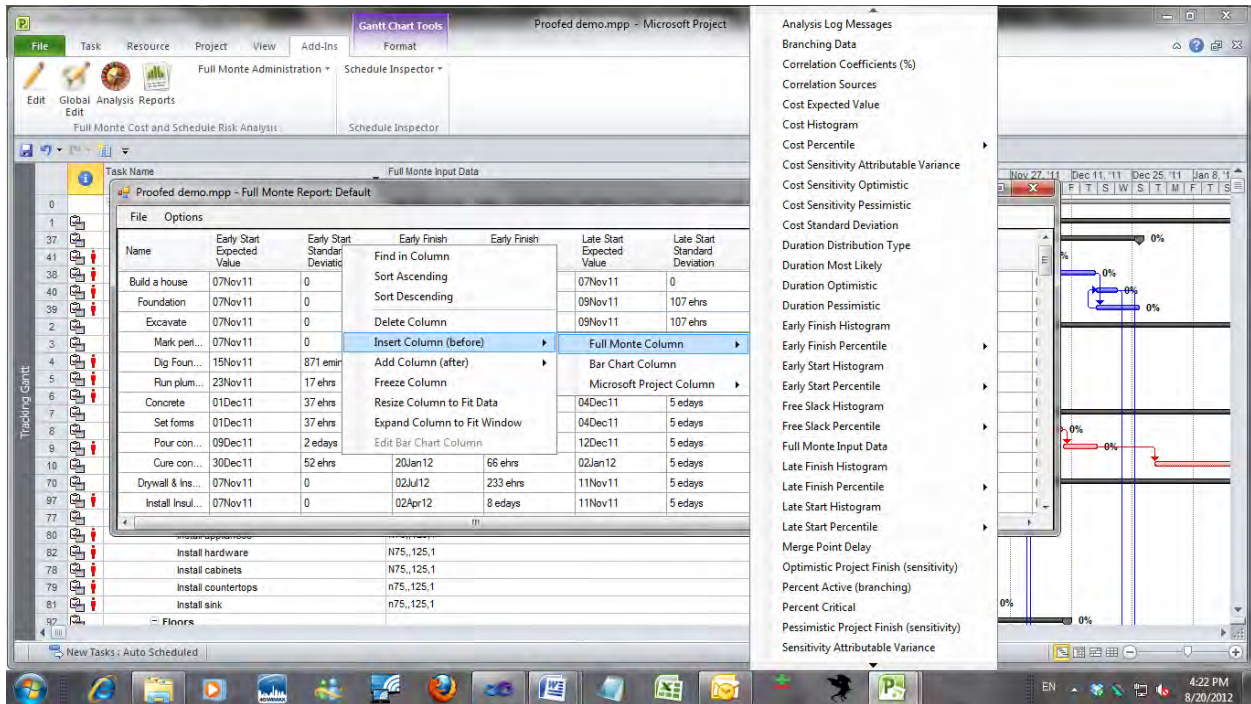
Figure 8.1

The Full Monte list includes all output fields, while the Project list includes only a small subset. Click on the field you want to include. The new column will appear after the column on which you right-clicked.

To insert a column: Follow the same instructions as to add a column, except select **Insert** rather than **Add**. The new column will appear before the column on which you right-clicked.

NOTE: You can also add a bar chart column. Bar charts are treated as a special kind of column in a tabular report. Full Monte offers several customization features when you Add or Edit a bar chart column, and these are described in the next subsection.

Freeze columns: Right-click on the rightmost column you wish to freeze. Select **Freeze Column**. A checkbox will appear next to this option indicating that you can unfreeze the column by clicking on the checkbox again. When you scroll horizontally, the column you froze and those to its left will not scroll.

NOTE: The checkbox also appears on all the columns to the left of the rightmost column you froze, which are also frozen. When you click again to unfreeze, it will unfreeze the column you right-clicked on and any unfrozen columns to its right. To cancel all freezing, you must right-click on the first —the leftmost — column.

Freezing columns also has an effect on the pagination of printed reports. If the report is too wide to fit on a single page, the frozen columns will be repeated on each horizontal continuation page.

Resize a column: There are three options for resizing columns:

- Drag the right edge of the column header

- Right-click on the column heading and select **Resize Column to Fit Data**

- Resize all columns by selecting the **Options** menu, then **Resize All Columns to Fit Data**

Change the order of the columns: Simply drag the column headers to where you want them to be.

Sort: To sort the rows based on the content of a particular column, click on the column header. The data will be sorted in ascending order of the data in that column. If you click again on the same column heading, this order will be reversed. You cannot sort on multiple columns.

Observe the task hierarchy: Select the **Options** menu, then **Observe Task Hierarchy**. A checkbox appears indicating that you can click again to reverse this change.

Task hierarchy refers to the parent-child relationships between summary tasks and lower-level tasks. Selecting this option has a twofold effect:

- If the Name field is displayed, it is indented like it is in Project

- Children of any given parent will be kept together during sorting, and sorted only within their sibling group. For example, when you sort on the name, you might see a report that looks like that in figure 8.2.

Project summary task: Select the **Options** menu, then **Include Project Summary**. A checkbox appears indicating that you can click again to reverse this change.

Empty rows: Select the **Options** menu, then **Allow Empty Rows**. A checkbox appears indicating that you can click again to reverse this change. (While a blank row has no significance, users of Project sometimes lay out their projects using blank rows for clarity.)

Show the Microsoft Project field names: Select the **Options** menu, then **Show Microsoft Project Field Names**. The Project field names will be appended in parentheses to the Full Monte field names in the column headers. A checkbox appears indicating that you can click again to reverse this change.
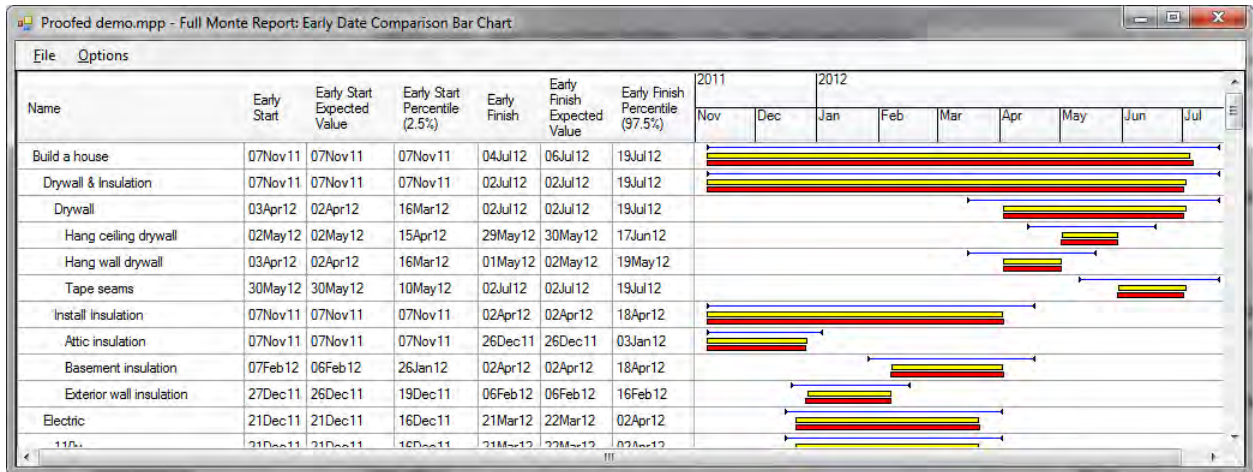
Figure 8.2

Date format: To change the date format, select the **Options** menu and click on the date format you want. A submenu will appear containing the current date and time displayed in various formats. Click on the one you would like to use. To read how to install more date format choices, see Appendix D, Date Formatting.

Bar charts

Add a Bar Chart Column: To add a bar chart column, right-click on a column header. Select **Add** or **Insert**, then **Bar Chart Column**. You'll see the Bar Definition dialog box, as shown in figure 8.3.
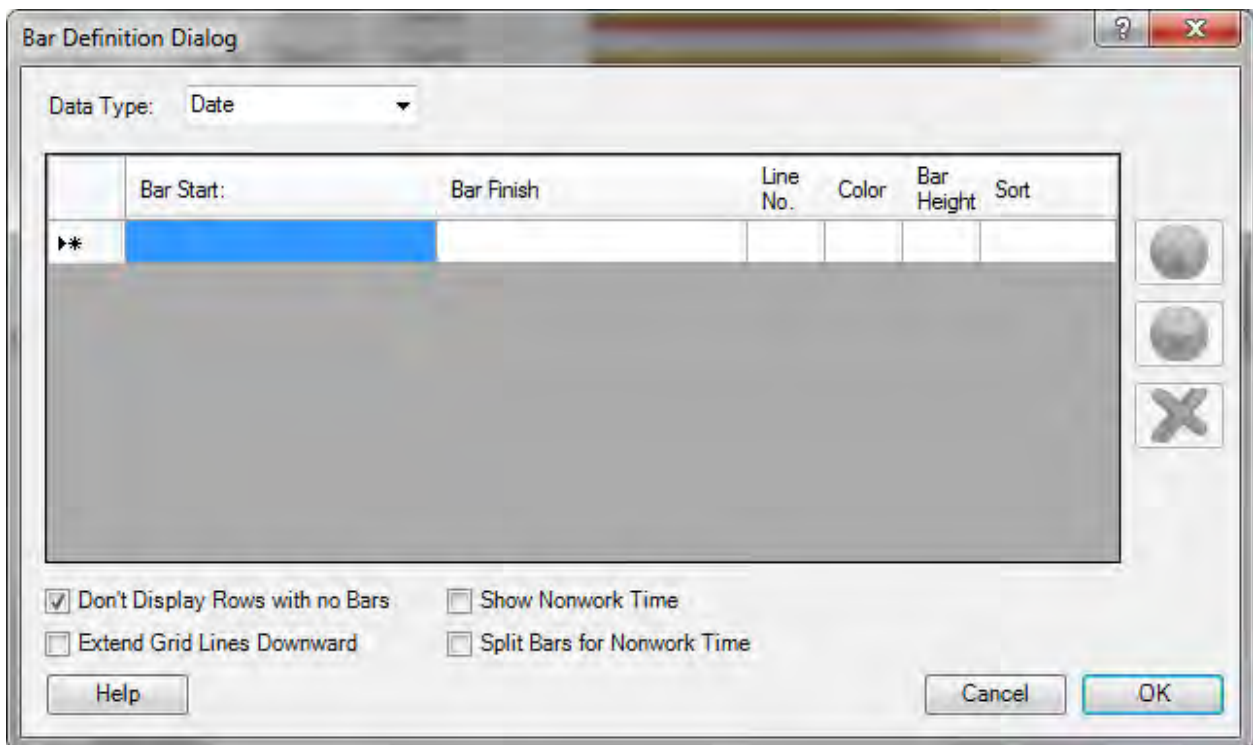
Figure 8.3

The drop-down menu at the top of the dialog box allows you to select the type of data to display. This must be date, duration, or cost/numeric. Since you cannot mix these types, once you have defined a bar this control is disabled.

In the grid, you can define a number of bars and/or milestone symbols. Bars extend from Bar Start to Bar Finish. For a milestone symbol, leave out one of the values. For example, for a finish milestone, leave out the From date. For a start milestone, leave out the To date. Select fields by selecting from the drop-down list that appears when you click on the cell or by right-clicking on the cell. For cost/numeric bars, there is an additional option for the Bar Start to be zero, enabling you to display bars with lengths that represent the value of a particular cost or numeric field.

Figure 8.4 shows a completed bar definition dialog.

Figure 8.4

Line Number: The line number allows you to separate bars and symbols vertically.

> NOTE: Bars are displayed in the order in which they are defined. Unless you know they won't overlap, you probably want to either separate them vertically or make the latter bar thinner. You can alter the display order using the up and down arrow buttons to the right of the grid.

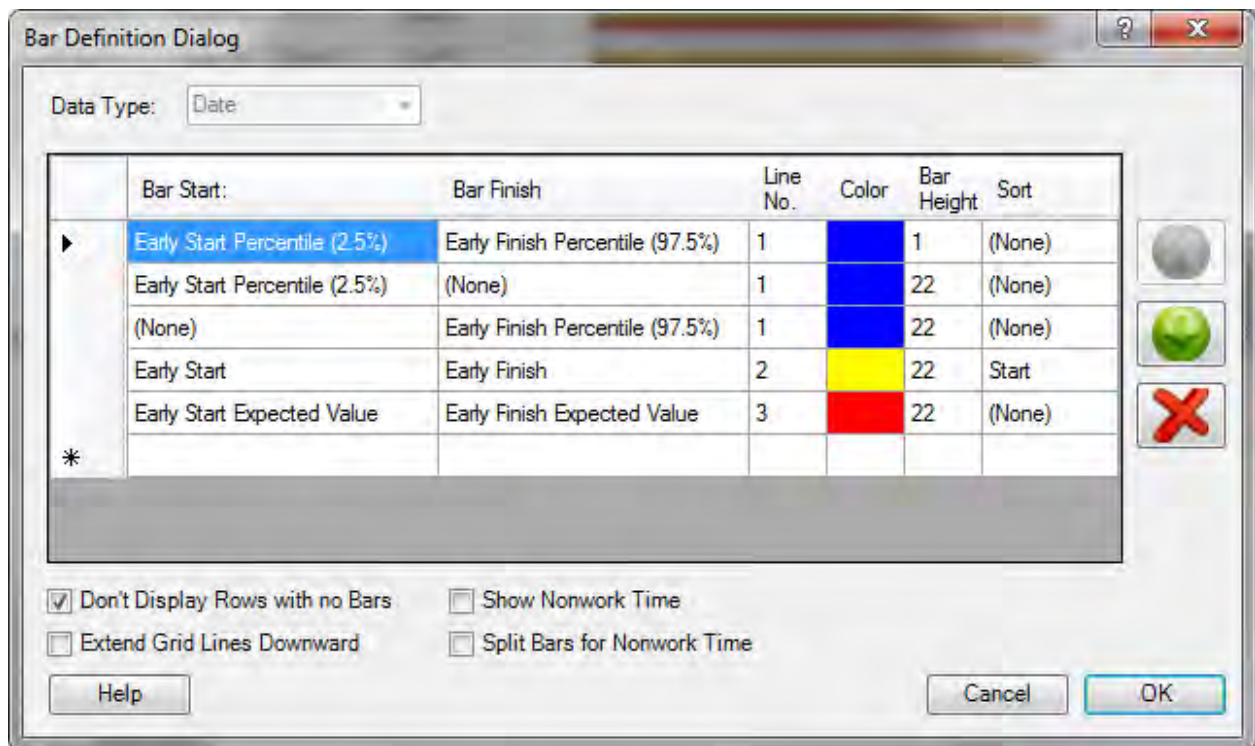Lines are numbered from top to bottom within the cell, and the numbers do not have to be consecutive (Full Monte closes up any gaps). If you don't enter a line number, it will be taken as 1.

>   *NOTE: A bar or symbol will not appear if the dates are not populated. For example, if you want to draw a bar from actual start to early start to show progress, it will appear only if the task has been progressed.*

Color: Clicking on the next cell invokes a color choice dialog box where you choose the color of the bar or symbol.

Bar Height: It is very important to enter the bar height, because a zero-height bar will simply not display.

>   *NOTE: If you want to temporarily suppress a bar without losing its definition, set its bar height at 0.*

The bar height is expressed as a percentage of the total cell height, and the heights are scaled if necessary, so that their total does not exceed 90%.

>   *NOTE: Full Monte applies a height to each line based upon the bar or symbol with the greatest height in that line, and then it applies the rule to the line heights. This avoids double counting bars and symbols on the same line.*

Defining Bars and Symbols: Enter a line for each bar or symbol you want to define. The grid will scroll if necessary.

Sort Order: The final column in the grid controls the sort order. (Note that it does not actually sort the report; it defines what will happen when you sort on this particular column by clicking on its header.)

Click on the cell and select the sort order from the drop-down menu at the top of the screen or by right-clicking on the cell. You can sort on only one bar or symbol, so when you select one the others will revert to (None). For *bars*, you can choose to sort on either end or on the length of the bar. For *symbols,* you can sort only on the date that it is based on.

There are four checkboxes controlling the display of the bar-chart column as a whole.

Don't Display Rows with no Bars will suppress any rows for which there are no bars. The main reason for there to be no bars is that the dates at both ends of the bar are the same for this particular task. For example, in the tornado report the bars appear only for tasks that can affect the project finish date, which is often a very small subset.

>   *NOTE: If there is more than one bar-chart column, this option has to be selected for all of them to have any effect. In order to suppress a row, there must be no bars on any of the columns.*

Extend Grid Lines Downward extends the lines from the lowest level of the date grid in the header row downward over the data rows.

> NOTE: This and the following two checkboxes are disabled if the data type is not Date.

Show Nonwork Time causes nonwork time to be displayed as a gray stripe underneath the bars (but seen next to the item). It displays these only to the granularity of the lowest-level grid markings so weekends will not show up unless the grid is marked in days or less.

Selecting Split Bars for Nonwork Time causes nonwork time to be indicated by a break in the bars. This break will be either white or gray depending upon the setting of the previous option. It splits the bars only to the granularity of the lowest-level grid markings, so weekends will not show up unless the grid is marked in days or less.

When you are done, click on **OK**. A column will appear before or after the place you right-clicked, looking similar to the bar chart shown in figure 8.5. In this example, the "extend grid lines downward" checkbox was checked.
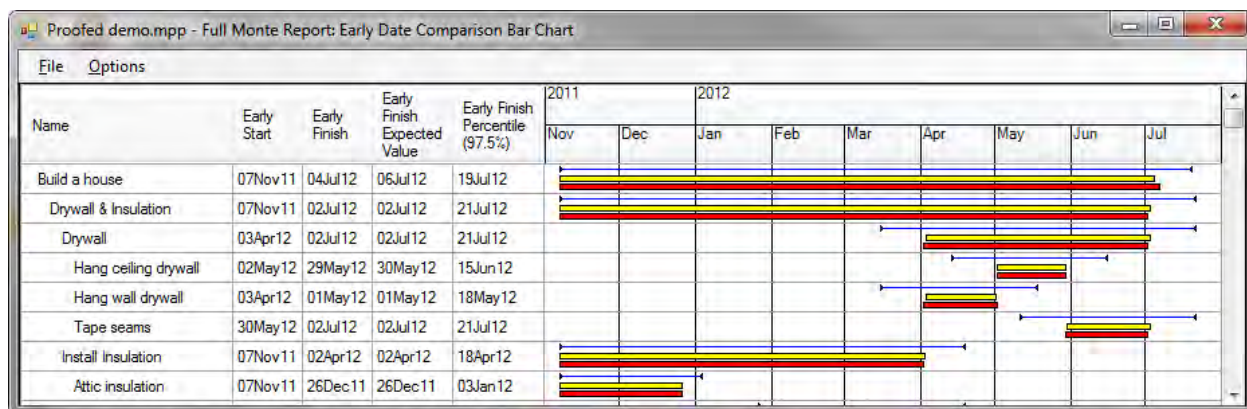
Figure 8.5

> NOTE: You can have any number of bar chart columns in a single report.

Header and footer in printed reports

Printed tabular reports and bar charts can have an optional header and/or footer on each printed page. This allows you to keep a record of things such as the date and time it was printed, the project, company, report names, and page numbers. The definition of the header and footer is stored with the report.

To customize the header and footer, open the tabular report. Select **File**, then **Edit Header/Footer** for a printed report from the menu. You should see the dialog box in figure 8.6.
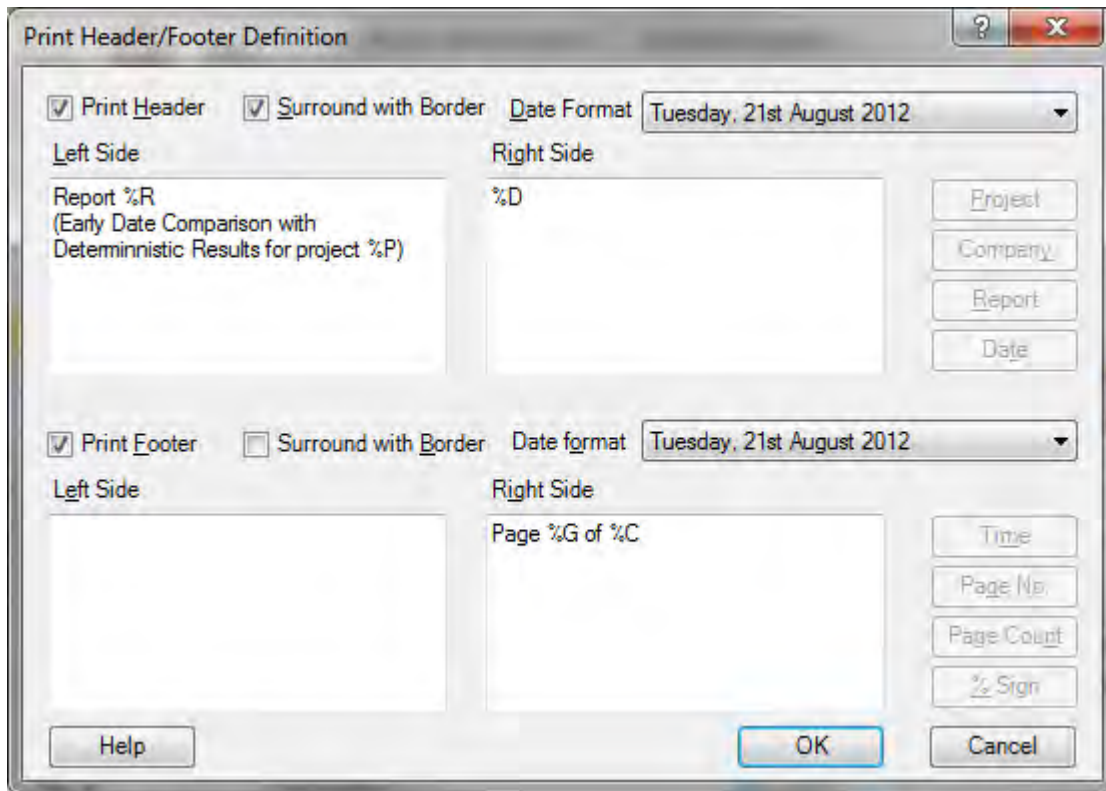
Figure 8.6

Checkboxes allow you to determine whether the header or footer is printed at all, and if so whether it is surrounded by a border. You can also select a date format for each, independent of the date format used in the body of the report.

The header section text is divided into two areas, which will be printed left and right justified, respectively, on the page. Note that the text to be displayed is defined in two separate boxes, though they do not appear on the report that way, and that the right-hand text is right justified in the printed report but not in the dialog box. Some text is simply reproduced on the report literally, while special values are defined by a % sign followed by a character. These get substituted for the appropriate values when the report is printed. Thus, "Page %G of %C" might appear as "Page 3 of 7."

You cannot actually type the special symbols into the text boxes; instead, position the cursors and then press the appropriate button.

The footer section works the same way. Figure 8.7 shows what the header and footer defined above would look like when printed.

Report Early Date Comparison Bar Chart   Tuesday, 21st August 2012
(Early Date Comparison with Determinnistic Results for project Proofed demo.mpp)
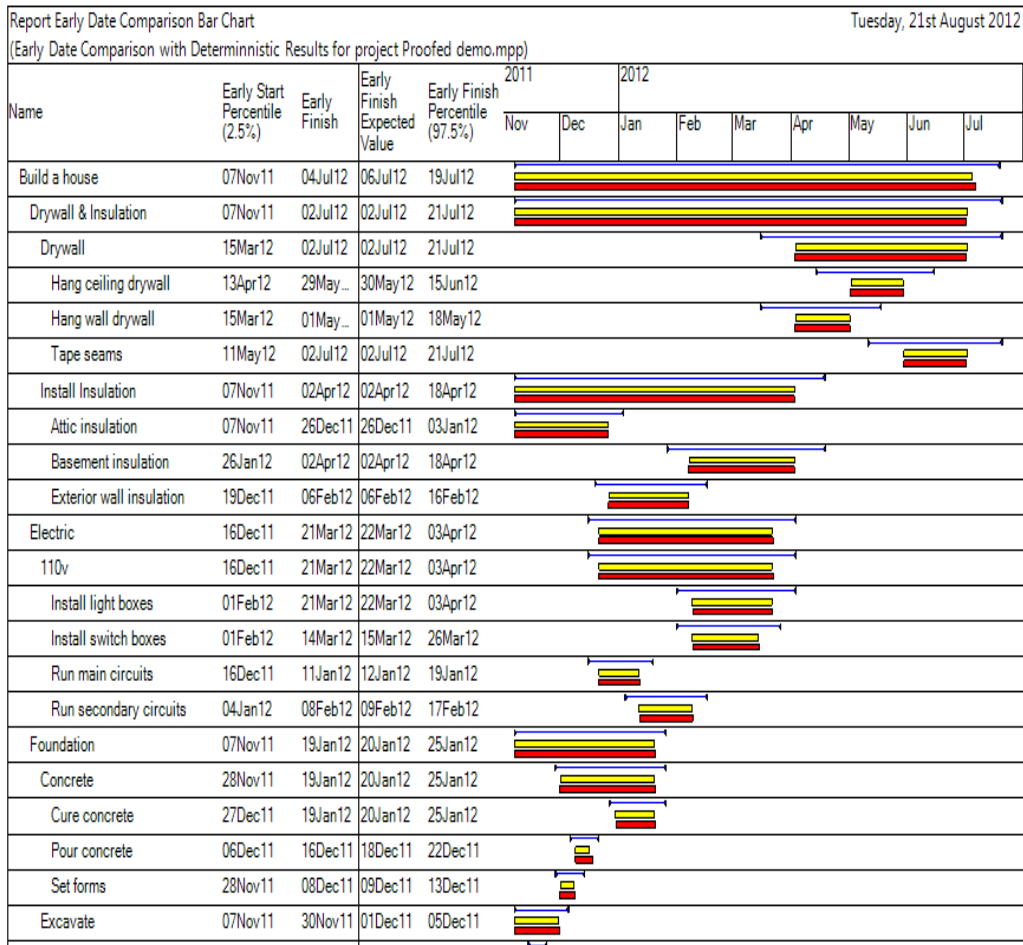
Name	Early Start Percentile (2.5%)	Early Finish	Early Finish Expected Value	Early Finish Percentile (97.5%)	2011 Nov	Dec	2012 Jan	Feb	Mar	Apr	May	Jun	Jul
Build a house	07Nov11	04Jul12	06Jul12	19Jul12									
Drywall & Insulation	07Nov11	02Jul12	02Jul12	21Jul12									
Drywall	15Mar12	02Jul12	02Jul12	21Jul12									
Hang ceiling drywall	13Apr12	29May…	30May12	15Jun12									
Hang wall drywall	15Mar12	01May…	01May12	18May12									
Tape seams	11May12	02Jul12	02Jul12	21Jul12									
Install Insulation	07Nov11	02Apr12	02Apr12	18Apr12									
Attic insulation	07Nov11	26Dec11	26Dec11	03Jan12									
Basement insulation	26Jan12	02Apr12	02Apr12	18Apr12									
Exterior wall insulation	19Dec11	06Feb12	06Feb12	16Feb12									
Electric	16Dec11	21Mar12	22Mar12	03Apr12									
110v	16Dec11	21Mar12	22Mar12	03Apr12									
Install light boxes	01Feb12	21Mar12	22Mar12	03Apr12									
Install switch boxes	01Feb12	14Mar12	15Mar12	26Mar12									
Run main circuits	16Dec11	11Jan12	12Jan12	19Jan12									
Run secondary circuits	04Jan12	08Feb12	09Feb12	17Feb12									
Foundation	07Nov11	19Jan12	20Jan12	25Jan12									
Concrete	28Nov11	19Jan12	20Jan12	25Jan12									
Cure concrete	27Dec11	19Jan12	20Jan12	25Jan12									
Pour concrete	06Dec11	16Dec11	18Dec11	22Dec11									
Set forms	28Nov11	08Dec11	09Dec11	13Dec11									
Excavate	07Nov11	30Nov11	01Dec11	05Dec11									

Figure 8.7

Graphical reports

Graphical reports — histograms and S-curves — can be accessed by using or creating a report that includes histogram columns. These show up as buttons giving access to the graphics. The customizations are stored with the report specification. To create any permanent customizations, start by opening the report you wish to customize. Then click on one of the graphics buttons.

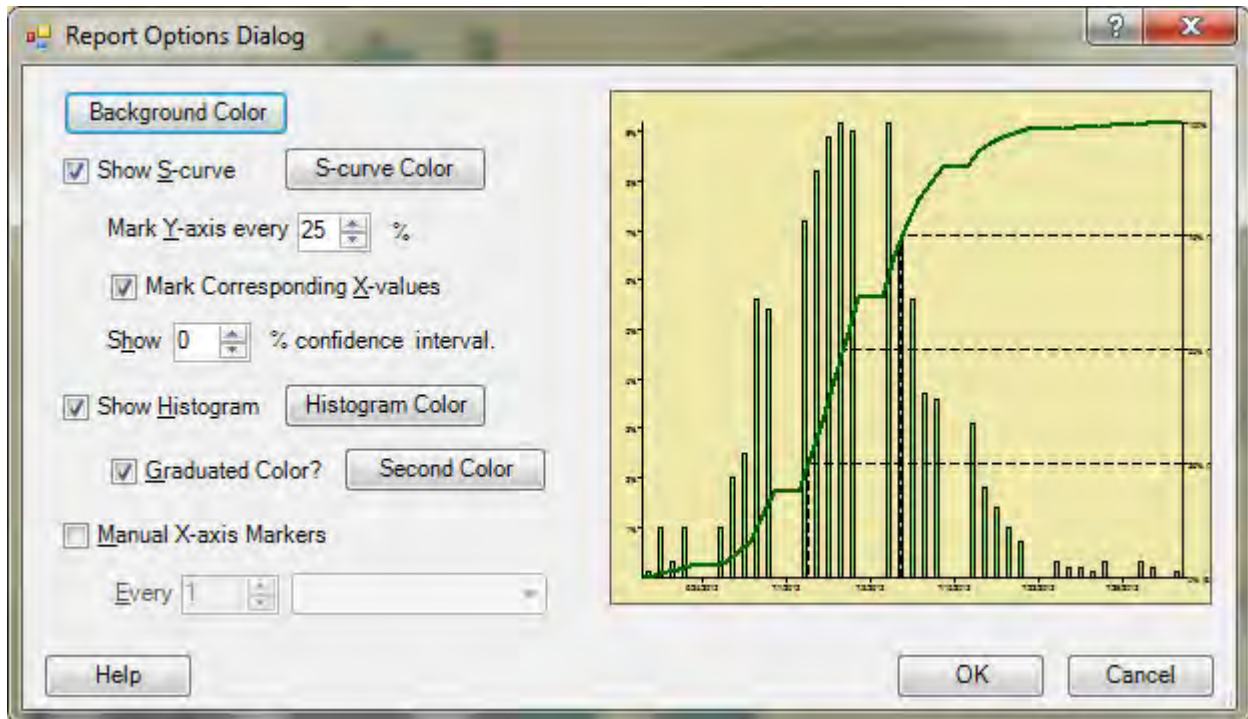Clicking on **Options** at the top of the window brings up the Graph Options dialog box, shown in figure 8.8.

Figure 8.8

Background Color: This button brings up a color selection dialog box from which you can select the color of the background. When you click **OK** on the color dialog box, the color you selected will be shown in the preview panel. It will not appear on the actual report until you click OK in the options dialog box.

NOTE: This is true for all the selections on this dialog box. Their effect is previewed in the dialog box but will not be made permanent until you click OK.

Show S-curve: Check this box to show the cumulative distribution function as an S-curve. The cumulative distribution function represents the probability that the actual value will fall to the left of a given point.

NOTE: You must select the histogram, the S-curve, or both. If you uncheck this box, Full Monte will automatically check the histogram box if it is not already checked.

If the above checkbox is checked, various options related to the S-curve will be enabled, as follows:

S-curve Color: This button brings up a color selection dialog box from which you can select the color of the S-curve. When you click **OK** on the color dialog box, the color you selected will be shown in the preview panel. It will not appear on the actual report until you click OK in the options dialog box.

Mark Y-axis: This allows you to determine the frequency of the Y-axis markings for the S-curve. The curve always goes from 0 to 100, and Full Monte always marks those two extreme points. While you would

normally choose a value that divides exactly into 100, this is not required. For example, if you set the interval to 90 it will mark 0, 90, and 100.

Mark Corresponding X-values: If you check this box, the report will show the value corresponding to the cumulative percentage on the S-curve in parentheses as part of the Y-axis marker.

Show Confidence Interval: By default, the S-curve is shown as a line, but this is only an estimate. How precise the estimate is depends upon the number of simulations performed. Increasing the confidence interval from 0 will replace the line with a pair of lines indicating the range of values between which the real line lies. The higher the confidence value, the wider the gap between the two lines. If you specify, for example, 90% and find the result unacceptable, then you should do more simulations.

Show Histogram: Check this box to show the frequency function as a histogram. The frequency function represents the probability that the actual value will fall within a given range represented by each bar.

> *NOTE: You must select the histogram, the S-curve, or both. If you uncheck this box, it will check the S-curve box if that is not already checked.*

If the above checkbox is checked, various options related to the histogram will be enabled, as follows:

Histogram Color: This button brings up a color selection dialog box from which you can select the primary color of the histogram bars. When you click **OK** on the color dialog box, the color you selected will be shown in the preview panel. It will not appear on the actual report until you click OK in the options dialog box.

Graduated Color: If you select this checkbox, the **Second Color** button will be enabled.

Second Color: This button brings up a color selection dialog box from which you can select a secondary color for the histogram bars. The color is graduated between the two and lends an attractive appearance for presentations, but has no other significance.

Manual X-axis Markers: Finally, you can opt to determine the frequency of the X-axis markers manually. This is generally not necessary because Full Monte usually makes sensible choices, but if you want to:

- Check the checkbox

- Enter a number

- Pick a time (or dollar) unit from the drop-down list

When you are done and are happy with the previewed results, click on **OK** to apply these to the report. The options you selected will be used in the current session until you change them. They will be saved with the report if you choose to save it.

CHAPTER 9: External Subprojects

Full Monte requires that the master project and all contained external subprojects have the same field mappings. While the easiest way to ensure this is to use the default mappings, this may not always be possible. When Full Monte accesses a project, it first checks the field mappings for the project as described in Chapter 2, Getting Started. If the project has any external subprojects, it then does further checks on the field mappings of these subprojects. If they match the mappings selected for the master project (either because these were already set or because they were set from the default values), then there is nothing else to do.

Mismatched field mappings

If, on the other hand, the mappings on a particular subproject do not match those on the master, then Full Monte may display one of two different dialog boxes, depending on whether the subproject has no mappings or has different mappings.

If there are no mappings specified for the subproject, Full Monte checks for collisions with existing data. If it finds any, it displays a dialog box such as this one:
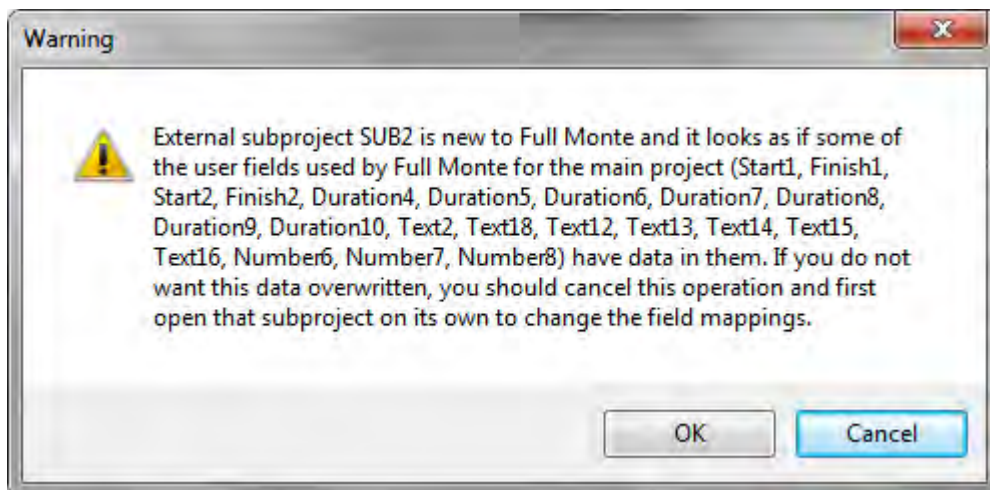

Figure 9.1

Unless you are certain that the data is unwanted, we recommend that you cancel the operation.

If, on the other hand, there are mappings but they are different from those in the master project, you will be given the opportunity to change the mappings to match those of the master project with the dialog box in figure 9.2:
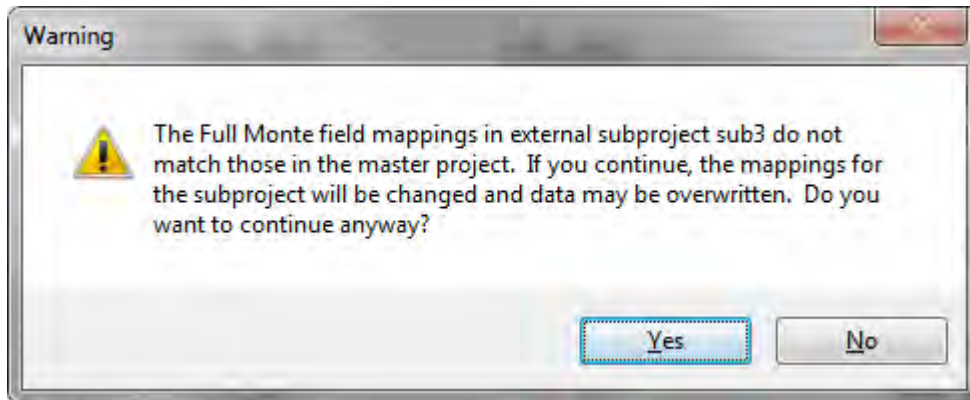
Figure 9.2

If you click on **Yes**, the mappings for the subproject will be changed and the existing data moved.

   NOTE: In this case Full Monte does not check for collisions.

APPENDIX A: Probability distributions

A probability distribution for a random variable is a function defining the probability of that variable having a particular value. (The distributions used in Full Monte are continuous, which means that the probability of any particular value is infinitesimal. It would be more accurate to talk about "the probability density at a particular value." Since Full Monte actually works in whole minutes and approximates the continuous distribution by a discrete one, we can set aside this nicety for now.)

Both the input to (task durations) and the output from (dates, slacks, and costs) Full Monte are probability distributions. The ones that come out are empirical. We don't have a formula for the function, we have an estimate based on simulation. Here we are concerned with the probability distributions we input for task durations.

Full Monte supports five families of probability distributions. Here we discuss what they are, how you choose among them, how the parameters are interpreted, and how Full Monte samples from them.

Uniform distribution

The simplest distribution type is the *uniform distribution*, in which any value in a specified range is equally likely. Unfortunately, it is also the least useful because it is unlikely to be realistic in practice. (If you think that a task will take between four and six weeks, the actual value is probably more likely to be toward the center of the range rather than near the outer edges.)

Sampling is merely a matter of translating a pseudorandom number (which emulates a true random variable distributed uniformly between 0 and 1) to the specified range. That range is specified by the user on the Full Monte Edit dialog box. Since the distribution is symmetrical, there is no need to input a middle value. (The dialog box calculates and displays what it calls the most likely value, though in this particular case it is actually the mean value. By definition, no value is more likely than any other.) An example of the uniform distribution as generated by Full Monte is illustrated in figure A.1.
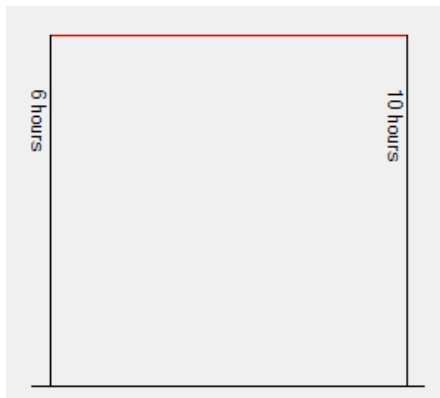

Figure A.1

All the other distribution types taper off to zero at their edges. Users are advised not to use the uniform distribution unless they have very good reason to expect the data to be uniform.

Normal distribution

The other symmetrical distribution supported by Full Monte is the *normal distribution*, which is much more useful. It follows a familiar bell-shaped curve and is often used to model a variety of natural phenomena. (Perhaps this is because if you add together a lot of random variables from *any* distribution, the result tends towards normal.)
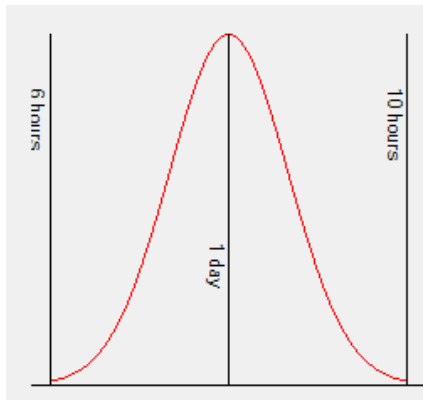
Figure A.2

Use this or the lognormal for most tasks unless you have a reason to expect the distribution to be asymmetrical. The remaining three distributions supported by Full Monte — lognormal, beta, and triangular — are asymmetrical, or "skewed."

Lognormal distribution

The *lognormal distribution* is skewed in a very special way. It is the distribution of a variable whose logarithm is distributed normally, which means it is defined by just two parameters. Therefore, you cannot specify the most likely value, so you cannot determine the skew independently of the optimistic and pessimistic values.

It is always skewed to the right, or "positively." The tail on the right is longer than the tail on the left. In fact, while it has an infinite tail to the right, its absolute minimum is always zero since the log of minus infinity is zero. When the optimistic and pessimistic values are close together the lognormal differs little from the normal, but as the ratio between the two increases the skew becomes greater.

The lognormal also occurs often in nature for values that can vary by orders of magnitude. (Which is why when you look at a list of populations of cities of all sizes, you find that about 30% of them have 1 as their first digit.) Use the lognormal for situations where the duration is highly uncertain, such as basic research projects, where the pessimistic estimate may exceed the optimistic by an order of magnitude

70

or more. There is even some empirical evidence from Dan Trietsch of the American University of Armenia that the lognormal distribution should be used in all cases. (See http://mba.tuck.dartmouth.edu/pss/Notes/LognormalParkinson.pdf.)

Two examples of the lognormal distribution as generated by Full Monte are illustrated in figure A.3. The first is less extreme, with the minimum value being 60% of the maximum, making it almost symmetrical, while the second has a minimum which is only 10% of the maximum.
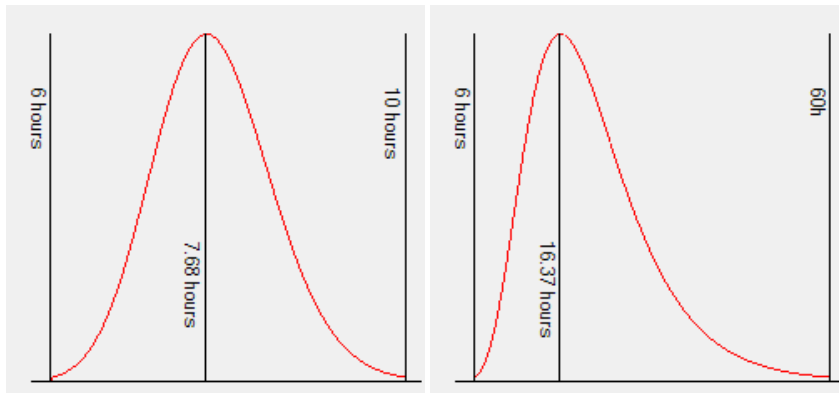
Figure A.3

The final two supported distributions are determined by three parameters and thus give the user complete control over the amount and direction of the skew.

Beta distribution

As interpreted by Full Monte, the *beta distribution* is the least theoretically defensible, but also may be the most useful distribution.

A true beta distribution is specified by no less than four parameters, which give rise to an astonishing variety of shapes, many of which would not be expected or useful, such as continuously increasing or decreasing or even a U-shape with the lowest point in the middle. It also includes the uniform distribution as a special case, but in Full Monte we treat this separately.

Full Monte, therefore, restricts the choice to a well-behaved subset that rises monotonically from zero at its most optimistic point until it reaches its peak, and then declines monotonically to zero at its most pessimistic point. In fact, the particular subset used by Full Monte when defined to be symmetrical looks almost identical to a similarly defined normal distribution and has an almost identical standard deviation. The standard deviation decreases, however, as the skew increases. Two examples of the beta distribution — one symmetrical and one skewed to the right — are illustrated in figure A.4.
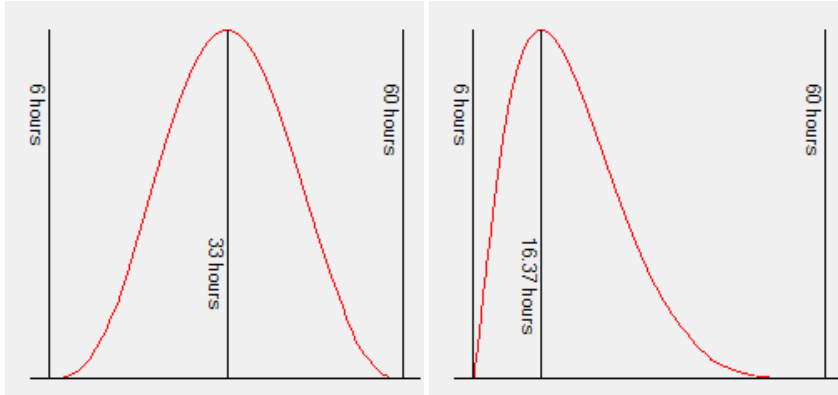
71

Figure A.4

Use this distribution when you want something similar to a normal distribution but want to control the skew.

Triangular distribution

Finally, the *triangular distribution* has long been used for schedule risk analysis. It does not appear much naturally, but it is simple, and because it puts more emphasis on values close to the limits, it is perhaps more in tune with users' expectations. As its name implies, it is simply a triangle with the base defined by the optimistic and pessimistic estimates, and the apex defined by the most likely value. It is very easy to sample and to understand, but like the uniform distribution, it is unlikely to reflect the real situation. Because its tails are not thin, it has a larger standard deviation than similarly defined distributions of all the other types except uniform. Two example of the triangular distribution — one symmetrical and one skewed extremely to the right with the same parameters as the beta above— are illustrated in figure A.5.
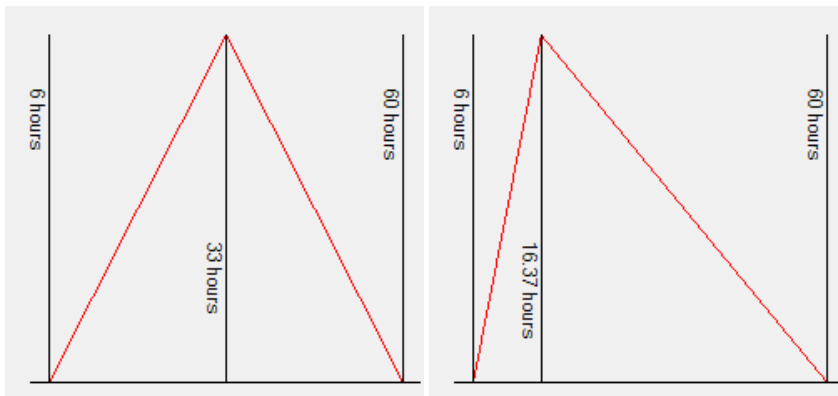
Figure A.5

APPENDIX B: Correlations

Correlation is the name given to the phenomenon of interdependence between random variables, such that the frequency function for one of them depends on the actual values of the others. For two variables, it is often illustrated visually by a scatter diagram. To fully describe the correlation between two or more variables mathematically, one must define their multivariate distribution function, which is a multidimensional extension of the frequency function defining the probability of each possible combination of values the variables. For most purposes, however, and certainly for input to Full Monte and other schedule risk analysis software, we summarize this using the concept of correlation coefficients.

A correlation coefficient measures the degree of interdependence between two random variables in terms of a number between -1 and 1. The details are beyond the scope of this guide and there are actually two different correlation coefficient definitions. The interested reader is referred to the URLs listed in the bibliography in Appendix F. Suffice it to say that a coefficient of zero means that the variables are independent, while a coefficient of + or - 1 means that they are perfectly (positively or negatively) correlated, i.e., the value of one variable exactly determines the value of the other.

It is worth noting, since it surprises some people, that perfect correlation can exist even between variables with different distributions, as illustrated in the scatter diagram in figure B.1. (Actually, the Pearson or linear correlation coefficient displayed is not quite 1 but the Spearman correlation coefficient is exactly 1. But more to the point, any particular value of one of the variables precisely determines the value of the other.)
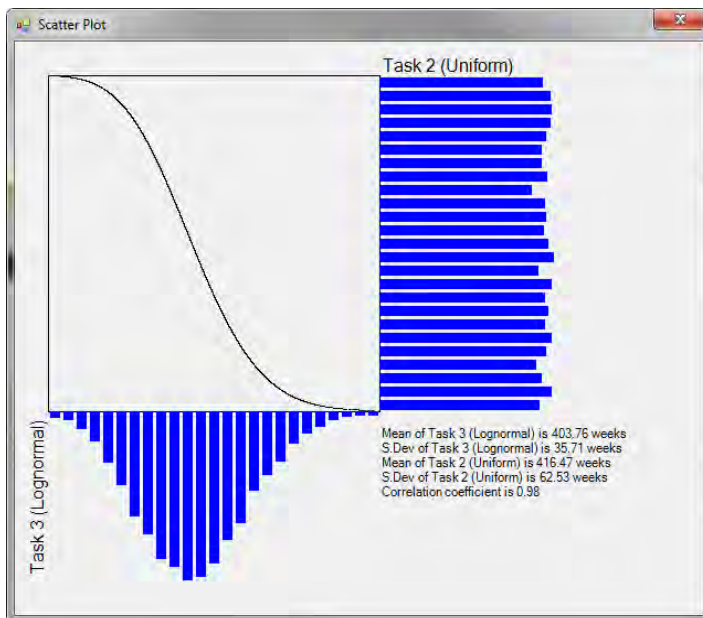

Figure B.1 100% correlation between a uniform and a lognormal distribution.

It is also worth noting that the correlation coefficient measures linear, or at least monotonic, relationships. A Wikipedia URL referenced in the bibliography in Appendix F shows how various

convoluted relationships can exist between variables that have zero correlation coefficients. The correlations in Full Monte can however be relied upon not to exhibit relationships and a typical correlation between two task durations is illustrated in figure B2.
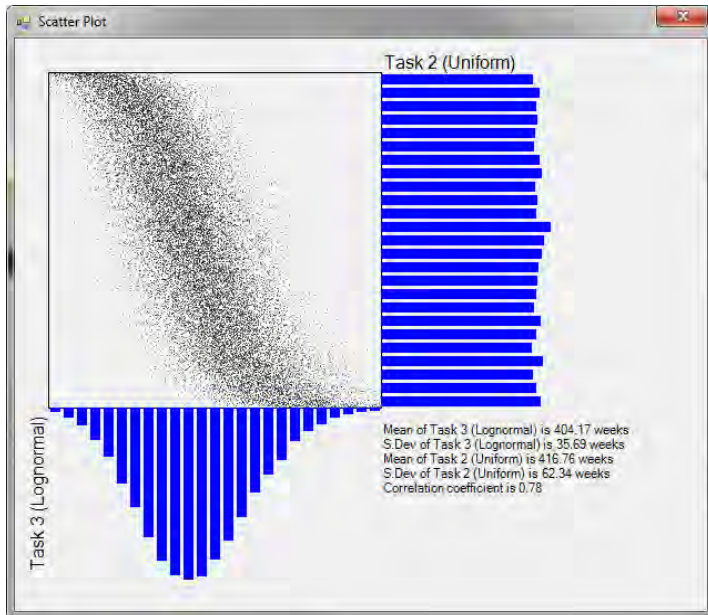

Figure B.2 Approximately 80% correlation between a uniform and a lognormal distribution.

So, correlation coefficients provide a shorthand way of defining the joint distribution of two variables, which more generally would require us to define the joint distribution in detail. But even using correlations coefficients, specifying the relationship between three or more variables can be a daunting task. To properly describe the correlations between the three tasks, one needs to define the correlations between each of the three possible pairings. And it gets much worse: If there are four tasks involved there are six possible pairings, and for five there are 10.

Full Monte handles correlations between the durations of tasks differently than some other systems. Most allow the user to specify the correlation between the durations of any two tasks, but this does not fully define the relationship between a larger set of tasks. Suppose that five tasks (A through E) are correlated in a symmetrical fashion, i.e., the correlations between any pair are the same. There are 10 possible pairings, but in simple systems only five places (one on each task) to enter a correlation with another task. What happens is that you pick a task arbitrarily (say A) and correlate the others with it. Let's say you put in 50% as the correlation coefficient. The result is that B, C, D, and E are correlated with A at 50%, but are correlated with each other only 25% (50% turned into a fraction and then squared). So the symmetry is not accurately modeled.

Some systems ignore this problem, which means they do not properly handle correlations involving more than two tasks, while others require the user to write code to set up the data in a strictly upper triangular matrix. Other systems support "drivers," an idea which seems superficially similar to that of

74

correlation sources, but does not really address this problem. These differences are discussed at the end of this appendix.)

Full Monte takes a different approach, recognizing that the usual reason for a correlation between the durations of different activities is that they all depend upon some outside factor, such as their location, which, for example, would make them all susceptible to the same weather. We make this explicit, so that however many tasks there are, each of them just needs one scalar quantity to define its correlation with a particular factor, and hence with all other tasks similarly defined. So all the user has to enter for each task is the casual factor (which can be given any arbitrary name like "weather" or "Hattie's Birthday Party") and the correlation coefficient. Any number of such factors can be included, and once specified for one task they appear on the drop-down list for the others.

Of course if the situation really is asymmetrical, so that tasks B and C really do depend on task A (and are therefore more correlated with A than with each other), then all you have to do is define a factor (say task A) and make A correlate with that with a coefficient of 100%.

Note that in the symmetrical situation described above, one should enter the square root of the correlation you want between each pair of tasks. So instead of putting in 50%, one would put in 70% (the square root of .5 is actually about 0.707).

Correlation Sources vs. Drivers

The driver concept is similar to the correlation source concept only in that they both allow the user to specify one or more outside factors to influence a set of task durations, but the way in which this influence is specified is very different. In the driver concept, the value of each driver is drawn from a distribution roughly centered on 1, and they are interpreted multiplicatively. For example, if a task is affected by two drivers these might be represented by two triangular distributions with parameters:

0 .7, 1.0, 1.2, *and*

0.9, 1.0, 1.4.

On a particular iteration, we might sample 1.04 and .98, so the value sampled for the task is multiplied by 1.04 times .98, or 1.0192.

While there is appealing logic to this seemingly simple idea, there are a number of problems for the unwary that can be overcome only with the most sophisticated analysis:

1. First, it means that the distribution specified for the task itself is not actually the distribution you get. Even in the above relatively simple scenario, most statisticians would be hard put to tell you what the final distribution is. In fact, one cannot readily calculate even the expected value or standard deviation. (They are about 1.06 and .15, respectively, by the way.)

75

2. Second, one cannot readily know what the correlation coefficient between two task durations actually is.

3. Third, you cannot easily specify that two tasks be affected by the same driver but to different degrees. (This is possible, but involves complicated calculations to get the right input data.)

4. Fourth, it is just additional data entry. It has been suggested that it can actually require less data entry, on the basis that all the uncertainty in task durations is attributable to drivers. Or that any remaining uncertainty can itself be attributed to a special driver. But this is a fallacy. Using the same driver to represent the uncertainty on two tasks would make them unintentionally correlated.

The net result is a system, which while it may have value in special circumstances, is likely to mean either much more work or unrealistic results.

By contrast, correlation sources require no extra data other than the name of the source. They enable the user to easily specify correlations between any number of tasks and any number of drivers and to simply calculate the correlation implied between any pair of task durations. (Full Monte even provides a visual aid in the form of a scatter diagram.) Furthermore, the distribution the user specifies for the task duration is unaffected by the correlation sources. One incidental advantage of this is that it is easy to assess the effect of correlations.

APPENDIX C: Branching

Branching is a feature that enables Full Monte to choose at run time, i.e. on each iteration of the simulation, which of a number of alternative paths from a task is active. Full Monte supports two types of branching: conditional and probabilistic.

In **conditional branching**, the user associates a date with each successor, such that the successor will be active if it can start before that date, and no other successor with an earlier associated date can.

In **probabilistic branching**, the user specifies a set of relative probabilities for the different successors, and the active successor will be chosen randomly based on these probabilities.

> NOTE: Either type of branching can be applied to the finish-to-start successors of any task that has two or more of these. It does not apply to other types of relationships, and you cannot combine the two different types of branching on successors to the same task.

It is important to understand that branching applies to a path rather than a task, which can lead to several rather confusing scenarios.

For example, if a successor task C is not on the selected active path out of task A, it could still itself be active if it has another predecessor B which is active. The general rule is that:

> A task is active if it is a start task (i.e., it has no predecessors, but see below about the task hierarchy) or if it has at least one active predecessor relationship from an active predecessor.

The situation is made more complicated by the task hierarchy in Project. A summary task is, in effect, both a start-to-start predecessor and a finish-to-finish successor to each of its children. For the purposes of branching, the start-to-start relationship is considered only if the child has no other predecessors, so it is in effect a "start task" within the summary task. So a child task is considered active as a result of its parent being active only if it has no other predecessors. The finish-to-finish relationships are never considered in this context, so a parent never becomes active purely as a result of one of its children being active.

Having stated the rules, it is probably best to avoid such confusing situations by using branching judiciously. We suggest that the use of branching be limited to situations where no successor task has any other predecessors apart from its parents.

Branching and progress

If any of the successors have been progressed, Full Monte assumes that the branch has been determined and will not apply the normal rules. In fact, it is quite possible (though illogical) that more

than one of the successors has been progressed. If any of the successors is progressed, Full Monte assumes that all the progressed successors are active and all the unprogressed ones are not.

Branching and the interpretation of calculated dates, slacks, and cost

Branching means, in general, that some activities are not always active. This is measured by the statistic "Active Percent," which estimates the percentage probability of each task being active. Some of the other calculated values are conditional upon the task being active, and some are not, as follows:

Mean and standard deviations of dates are conditional because it is impossible to assign a date value to something which did not happen. The expected finish date reported is the expected finish date, *given that the task is active.*

Mean and standard deviations of slacks are also conditional to make them consistent with the dates and because assigning a value of zero to the slack of a task that did not happen is not very meaningful. (Zero float means the task is critical, yet how can a nonexistent task be critical?) The expected total slack reported is the expected slack *given that the task is active.*

Mean and standard deviations of costs are, on the other hand, unconditional. This is reasonable because assigning zero cost to a task that did not happen makes sense. It is also necessary if the costs are to roll up correctly; if two alternative tasks with equal probability have an expected cost of $10,000 and $20,000, respectively, then the expected cost of the pair is $15,000, not $30,000.)

Criticality and sensitivity are also unconditional, though this choice is somewhat arbitrary. The value reported for criticality is the percentage of times overall that the task was critical, including the times when it was not critical because it was not active. (One can easily calculate the conditional criticality or sensitivity by dividing by the active percentage.)

Histograms (probability density) and **S-curves** (cumulative probability) are all conditional, which, in the case of cost, is somewhat inconsistent with the above. Doing otherwise would mean a tall bar at the origin representing the times when the cost was zero because the task was not performed, which would severely reduce the precision visible for the interesting part of the graph. It does mean, however, that the expected cost reported (which is *not* conditional) may be nowhere near the middle of the graphed data.

APPENDIX D: Date Formatting

You can change the date format in a report by selecting Date Format from the Options menu, which displays the current date and time in a selection of formats, as shown in figure D.1.
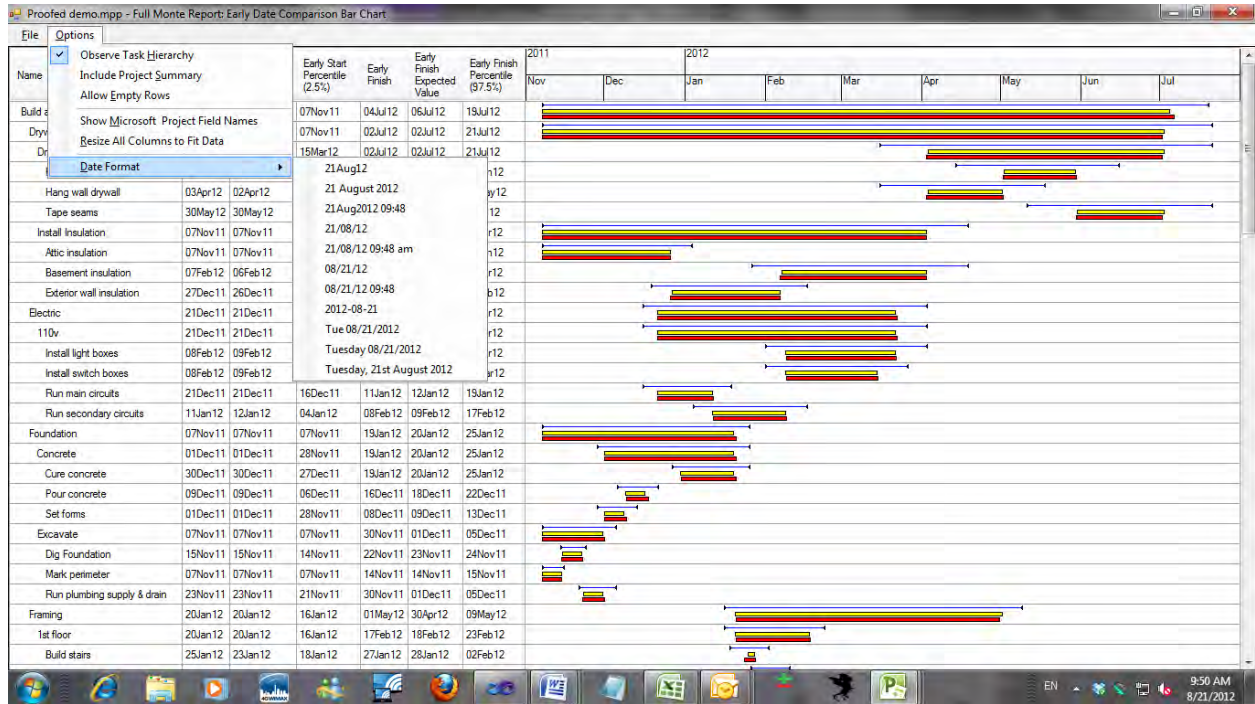

Figure D.1

This is only a small subset of the date formats Full Monte is capable of. The formats shown are controlled by a file called DateFormats.txt on the folder Barbecana/Full Monte below your "My Documents" folder. This can be edited using a simple editor like Notepad.

For example, the file that created the formatted dates in the drop-down menu above looks like this:

```
VDay
VD l Y
VDaY H:T
VD/M/y
VD/M/y H:T p
VM/D/y
VM/D/y H:T
VY-M-D
Vw M/D/Y
VW M/D/Y
W, o l Y
```

Apart from the punctuation characters, each character has a special meaning as follows:

A displays a short (3-character) abbreviation of the month in upper case.

a displays a short (3-character) abbreviation of the month in mixed case.

D displays the day of the month always in 2 digits (i.e. with leading zero if necessary).

d displays the day of the month in 1 or 2 digits (i.e. without leading zero).

H displays the hour of the day always in 2 digits (i.e. with leading zero if necessary).

h displays the hour of the day in 1 or 2 digits (i.e. without leading zero).

L displays the full name of the month in upper case.

l displays the full name of the month in mixed case.

M displays the month as a numeric always in 2 digits (i.e. with leading zero if necessary).

m displays the month as a numeric in 1 or 2 digits (i.e. without leading zero).

O displays the day of the month always in 2 digits and followed by "st," "nd," "rd"or "th."

o displays the month as a numeric in 1 or 2 digits and followed by "st," "nd," "rd" or "th."

P displays "AM" or "PM." (See also note below.)

p displays "am" or "pm." (See also note below.)

Q displays the quarter of the year in 1 digit.

T displays the minute of the hour always in 2 digits (i.e. with leading zero if necessary).

t displays the minute of the hour in 1 or 2 digits (i.e. without leading zero).

V anywhere in the string is not displayed directly but causes validation as explained below.

W displays the full name of the day of the week.

w displays a 3-character abbreviation of the day of the week.

Y displays the year in 4 digits.

y displays the year in 2 digits.

Notes:

1. If the time is displayed but there is no "p" or "P" in the format, the time is displayed using the 24-hour clock.

2. If there is a V in the format, Full Monte checks that the date is valid in Project (which supports only dates between 1984 and 2049) and if not displays "NA."

3. All other characters in the format, including spaces, are rendered as is.

APPENDIX E: Full Monte Output Fields

Full Monte reports on seven different items for each task:

1. Early Start

2. Early Finish

3. Late Start

4. Late Finish

5. Total Slack

6. Free Slack

7. Cost

Each of these has the same meaning as it does in Project, except that instead of being represented by a single value, each of the above is a probability distribution represented by three different fields:

1. The expected value

2. The standard deviation

3. The histogram/S-curve data

The histogram/S-curve data is stored in custom text fields in a format that is not designed for direct viewing but must be interpreted by Full Monte. These fields are also used to calculate the percentiles that are available in Full Monte reports.

The above accounts for 21 fields and these are named so that fields relating to the same basic data are sorted consecutively. Thus, for example, there are fields named:

• Early Start, expected value

• Early Start, standard deviation

• Early Start, histogram

An additional 12 fields do not fit into the above pattern. They are, in alphabetical order:

Analysis Log Messages: This field contains messages generated during risk analysis and relates to issues with each task. They can be viewed in the Message Log report, which can be shown automatically after risk analysis.

Cost Sensitivity, Attributable Variance: This is the percentage of the variance in the cost of the "cost sensitivity target," which is attributable to the current task. (The cost sensitivity target is a summary task nominated by the user. By default, it is the project summary task, representing the whole project. See Chapter 4, Analyzing the Project.)

Cost Sensitivity, Optimistic: This is an estimate of the expected cost of the cost sensitivity target if the current task's cost were to take its optimistic value.

Cost Sensitivity, Pessimistic: This is an estimate of the expected cost of the cost sensitivity target if the current task's cost were to take its pessimistic value.

Full Monte Input Data: This is included for completeness, but it is actually an input rather than an output from the risk analysis process. All user input data on duration distributions, correlations, and branching is stored in this field in a compact form not intended to be viewed directly. Most of this information is presented as separate "fields" in Full Monte reports.

Merge Point Delay: When a task has multiple predecessors, it cannot start until all of these predecessors are complete. If two or more of these predecessors are sometimes critical, there will be a gap between the latest of the expected early finishes of the predecessors and the expected early start of the successor, due to the phenomenon known as merge bias. (See the "Merge Bias" section in Chapter 1, Introduction.) This field measures the size of this gap.

Percent Active: This field merely measures the percentage of trials when the task was active. (Inactivity is caused by branching, as explained in Appendix C.)

Percent Critical: This field merely measures the percentage of trials when the task was critical.

Sensitivity Index: This field is intended to give a rough guide of how sensitive the early finish of the schedule sensitivity target is to variation in the task's duration. It is computed as the percent critical multiplied by the standard deviation of the duration and divided by the standard deviation of the early finish of the sensitivity target.

Schedule Sensitivity, Attributable Variance: This is the percentage of the variance in the early finish of the "schedule sensitivity target," which is attributable to the current task. (The schedule sensitivity target is a milestone task nominated by the user. By default, it is the project summary task representing the whole project. See Chapter 4, Analyzing the Project, and Chapter 6, Investigating Task Sensitivity.)

Schedule Sensitivity, Optimistic: This is an estimate of the expected early finish date of the schedule sensitivity target if the current task's duration were to take its most optimistic value.

Schedule Sensitivity, Pessimistic: This is an estimate of the expected early finish date of the schedule sensitivity target if the current task's duration were to take its most pessimistic value.

APPENDIX F: Bibliography

Hoel, Paul G., *Introduction to Mathematical Statistics,* Wiley, 1984. An academic introductory text on the subject of probability.

Savage, Sam L. *The Flaw of Averages:* Why We Underestimate Risk in the Face of Uncertainty, Wiley, 2012. A readable popular account of why basing decisions on deterministic average values can be dangerously wrong.

Schiller, John, et al. *Schaum's Outline of Probability and Statistics,* Schaum, 2008. The "clip notes" version.

URLs:

Empirical arguments for the use of the lognormal distribution:

http://mba.tuck.dartmouth.edu/pss/Notes/LognormalParkinson.pdf and
http://mba.tuck.dartmouth.edu/pss/Notes/PERT21.pdf

Full range of the beta distribution:

http://en.wikipedia.org/wiki/Beta_distribution

Scatter pots of variables with different correlation coefficients, including some which have a coefficient of zero but which are clearly not independent:

http://en.wikipedia.org/wiki/Correlation_and_dependence

For the math behind the two-dice analogy of merge bias:

http://barbecana.com/merge-bias/ and download the spreadsheet.

INDEX