Global Knowledge®

Expert Reference Series of White Papers

A Toolkit for Project Time Estimation

A Toolkit for Project Time Estimation

Perry McLeod, CBAP, PMP, PBA, SMC

Introduction

Time estimation on even the simplest project is not as easy as it seems. Many business analysts (BAs), project managers (PMs), and other project team members, such as subject matter experts (SMEs), customers, users, and other stakeholders work in a functionally organized environment. Functional settings force us to bring project team members together from other areas within the organization.

This organizational paradigm causes many issues, such as competing for resources, self-interested management behavior, poor coordination between projects, over commitment of resources, and a fundamental disregard for best practice project planning techniques. In the case of opportunistic management behavior, many functional organizations allocate resources based on project priority. In such cases, there is an incentive for project sponsors and senior managers to keep priorities high by any means possible. On the other hand, those who already have resources assigned to their projects would want to protect them from poaching. As we may expect, this poor time management behavior leads to a negative effect on project accounting practices.

Organizations often account for costs based on hours spent by team members on projects. In contrast, time devoted to internal activities, such as meetings, is viewed as a non-project expense. As a result, there is a built-in incentive for management to keep as many people as possible working on projects. A side effect of this is the lack of availability of resources for new projects. Moreover, real project costs are never identified and meetings are not tracked.

If our requirements team was at our disposal 100 percent of the time, always completed activities on target, and worked a full eight-hour day without distraction or a loss of productivity, estimating time would be a mere 1:1 ratio as seen in Equation 1.

$$d \equiv e$$

Equation 1. Effort Equals Duration

In this example, duration (d), the terminal calendar period it takes from the time the work begins to the moment it is completed, and effort (e), the actual amount of work required to accomplish the task (usually measured in hours), are identical ($\equiv$). Stakeholders are seldom at our disposal, no one can work all the time, and there are always other duties to perform.

In this paper, we will understand the real cost of multitasking. Next, we will explore standard approaches to estimation and alternatives to a single-point estimate such as the requirements elicitation, planning, analysis, and collaboration (REPAC) technique. We will also consider how elements such as productivity, cumulative probabilities and probability densities, and Bayesian reasoning can help us account for the uncertainty given in estimates. Lastly, learn how to determine the overall likelihood of task completion by the desired date.

The Problem

Planning project timelines is not absolute. Estimates based on probabilities must always come with a margin of error. Figure 1 illustrates the best practice order of magnitude with respect to project time and money. (This is known as The Cone of Uncertainty). Classic time estimation on projects done using two simple techniques yields a broad range of results. The first method, as seen in Equation 1, assumes that the person giving the estimate can account for the variables needed to provide a thoughtful assessment. Regrettably, this approach has a less than 30 percent success rate (Standish Group, Chaos Report, 2012) and is, ironically, used in most cases—even when the project timelines carry substantial risk.

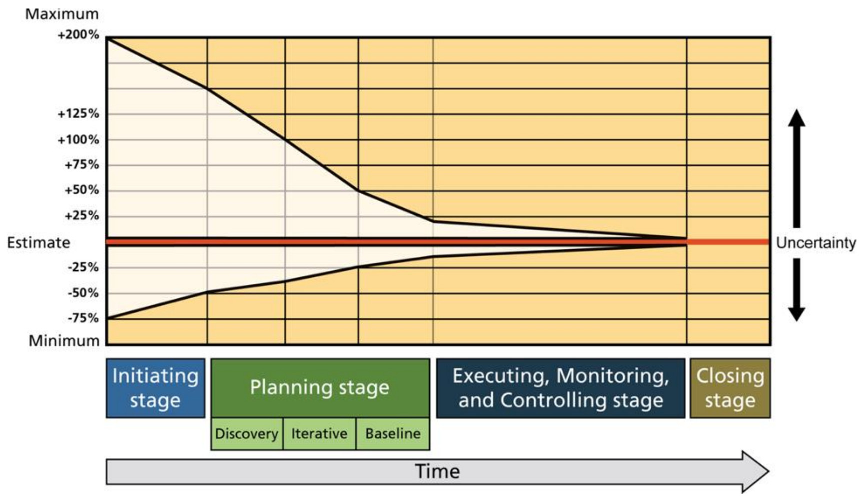

Figure 1. Project Margin of Error for Time and Effort (Global Knowledge, 2013)

It is not reasonable to expect project members and stakeholders to assess the inconsistency of their work forecasts. As a species, we have no intuitive sense of the variation of an uncertain quantity of time. The second classic technique accounts for stakeholder availability but still only offers a single point estimate of work time. Equation 2 is an example of availability as a consideration.

$$d = \frac{e}{a}$$

Equation 2. Single Point Estimate with Availability as a Variable

In this example, the duration (d) is equal to the amount of effort required to accomplish the task divided by the availability (a) the stakeholder has to work on the task. We cannot assume that our resource will work on the deliverable 100 percent of the time. Ideally, resources should dedicate themselves to a task 100 percent of the time that they are working on that task. Culturally, however, there is the expectation that project team members and stakeholders multitask. Rather than trying to change this behavior, it is far simpler to account for it mathematically and report this justification to project team members. Equation 3 illustrates how availability can dramatically affect project deliverables.

$$d = \frac{10\ hours}{50\%} = 20\ hours$$

Equation 3. Single Point Estimate with Availability as a Variable Example

Equation 3 is telling in that the projected ten hours of work has doubled because the resource is only available to work on it for half of their time. Availability for most projects is not a constant and tends to change frequently due to work habits, such as multitasking. Each time we switch from one task to another, time is required to bring ourselves back to the productive state we were in previously. Recent estimates indicate we can lose up to 40 percent of our productivity if we multitask (Weinschenk, 2012).

The Cost

Estimating time in projects is typically done using a single point estimate derived from experience and best guess. As mentioned in the introduction, many organizations are a functional chain of command and control. The structure is hierarchical in nature wherein people are grouped together according to their area of specialization, title, department, knowledge domains, and other means. This all too familiar arrangement demands that stakeholders and project staff take on multiple roles and multitask deliverables as they stretch themselves over many projects. Paradoxically, this is not something our brains can do.[1] Multitasking is a misnomer; the term *task-switching* more accurately describes this phenomenon. People who consider themselves "multitaskers" would be more precise to mention their ability to switch from task to task.

Those who make this boast may be fooling themselves. With the exception of walking and talking (something our brains have built extensive patterns for), task-switching uses up an enormous amount of cognitive energy and involves no fewer than four major areas of our brain. The switch itself may seem quick—hence the illusion. However, the ability to become productive at the task, especially when we are picking it up from a previous state, can be very time-consuming, emotionally draining, and cognitively expensive. Even so, we juggle multiple assignments, estimate when they will be complete, and try to keep timelines that never had a chance in the first place. To visualize this, let us assume a scenario wherein a deliverable on a project's critical path is given to a Business Analyst who already has other tasks on her plate. To deliver on time, she must manage this task against her other assignments. Tables 1 and 2 and Figure 2 provide some insight into how time slips away from the BA over the course of the day.

Table 1. Time Slippage over a Day

Time	Task	Minutes Elapsed	Minutes Worked	Minutes Not Worked	Productivity	Minutes Remaining	Percent Remaining
9:00:00 AM - 9:45	Main	45	30	15	67%	180	86%
10:00:00 AM - 10:15	Task 1	15	10	5	67%	50	83%
10:30:00 AM - 10:45				Morning Break			
11:00:00 AM - 11:30	Task 1	30	27	3	90%	23	38%
11:45:00 AM - 12:00	Task 3	15	7	8	47%	23	77%
12:15:00 PM - 1:00				Lunch Break			
1:15:00 PM - 1:45	Task 3	30	23	7	77%	0	0%
2:00:00 PM - 2:45	Main	45	35	10	78%	145	69%
3:00:00 PM - 3:15	Main	15	10	5	67%	135	64%
3:30:00 PM - 5:00	Main	90	75	15	83%	60	29%
5:15:00 PM - 6:00	Task 1	45	23	22	51%	0	0%
6:15:00 PM - 7:45	Main	90	60	30	67%	0	0%

Table 2. Time Slippage over a Day Summary

Total Work Minutes	Actual Work Minutes	Total Minutes Tasked	Total Minutes Not Tasked	Total Minutes Worked	Total Minutes Not Worked	Average Downtime	Average Productivity
480	390	330	60	132	48	12	71%

[1] I am referring to tasks where significant cognition is required.

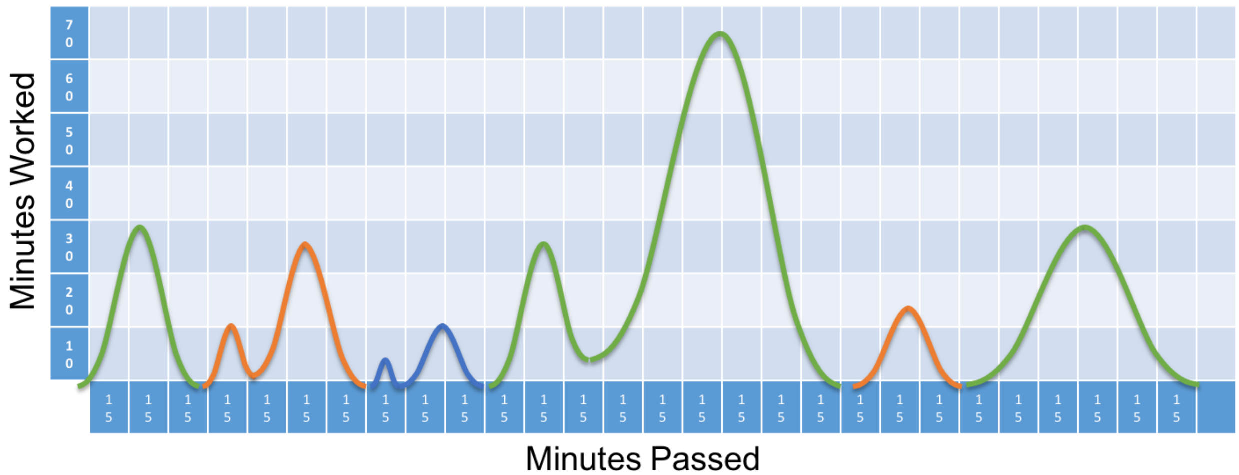

Figure 2. Time Slippage over a Day Distribution Curves

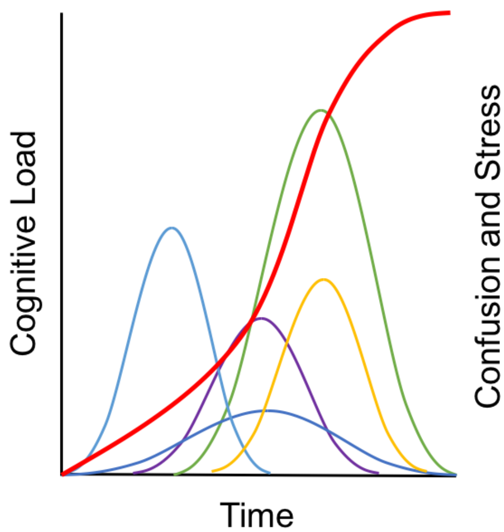

Figure 3. Overlapping Tasks Blur Cognition

As you can see, Table 1 skips between the main task and Tasks 1 and 3. What happened to Task 2? It completely fell of the BA's plate. Although we have the perception that we have completely moved from one task to another, the reality is more indicative of Figure 3. Here we see the thought patterns for each task overlapping. Each curve in Figure 3 represents a task, which corresponds to Figure 2. As you can see, there are more colors represented in Figure 3. This illustrates other distractions that compounded the BA's stress and confusion. These illustrations provide insight on how time slowly slips away from us over the course of one day. Overall, the BA was only productive for about 71 percent of the day. Known as the fifteen-fifteen effect, a person is only effective for about 70 percent of the day. Fifteen percent of the day is typically taken up by non-project-related distractions while the other 15 percent is attributed to personal matters, such as getting coffee, phone calls, web surfing, and emails. Compound this by several days or weeks, and we can see why deliverables are often late. We lack the capacity to imagine with any accuracy the length of time something will take as well as the ability to juggle multiple deliverables.

The Solution

For many years, science has proven that we can only do one thing at a time. More specifically, we can attend to only one cognitive task and process one and only one mental activity at a time: we can either talk or read but not do both at the same time. We can only have one thought at a time, and the more we force ourselves to switch from one thing to another, the more we tax our mental faculties. In experiments published in 2001, Joshua Rubinstein, Ph.D., Jeffrey Evans, Ph.D., and David Meyer, Ph.D., conducted four experiments in which young adults switched between different tasks, such as solving math problems or classifying geometric objects. For all tasks, the participants lost time when they had to switch from one task to another. As tasks became more complex, the participants performing them lost more time. As a result, people took significantly longer to switch between increasingly complex tasks. Time costs were also greater when the participants switched to tasks that were relatively unfamiliar. According to Meyer, Evans, and Rubinstein,

> Converging evidence suggests that the human "executive control" processes have two distinct, complementary stages. We call one stage "goal shifting" ("I want to do this now instead of that") and the other stage "rule activation" ("I'm turning off the rules for that and turning on the rules for this"). Both of these stages help people to, without awareness, switch between tasks.
>
> Although switch costs may be relatively small, sometimes just a few tenths of a second per switch, they can add up to large amounts when people switch repeatedly back and forth between tasks. Thus, multitasking may seem efficient on the surface but may actually take more time in the end and involve more error.

Meyer has said that even brief mental blocks created by shifting between tasks can cost as much as 40 percent of someone's productive time. It seems unlikely that we will be able to refocus company culture to accept the virtues of scheduling and completing one task before starting another. In another study from the University of California at Irvine, researchers shadowed workers on the job and studied their productivity. Study lead Gloria Mark revealed the following:

> You have to completely shift your thinking, it takes you a while to get into it and it takes you a while to get back and remember where you were . . . We found about 82 percent of all interrupted work is resumed on the same day. But, here's the bad news — it takes an average of 23 minutes and 15 seconds to get back to the task.

It is important to pull ourselves away from work now and then; in fact I teach my students a rhythm of 25 minutes to task, followed by a five-minute break, followed by another 25 minutes (repeat until the task is complete). Breaks are one thing, but distractions are another. Breaks are short, focused, and deliberate. Distractions catch us off guard and derail our task entirely. If this behavior is inevitable and unpredictable, what can we do about it? We could try a single point estimate using the resources' availably and productivity.

Single Point Resource Availablity and Productivity Technique

We have already seen that estimating time in projects is typically done using a single point estimate derived from experience and best guess. At best, a single point estimate gives us a 50 percent probability of success. We can increase those odds a few percentages by accounting for both the availability of a resource and their average productivity, which we have learned through studies is between 72 to 74 percent. I will use a constant of 70% for simplicity. In Equation 4 that follows, d represents duration, e represents the amount of effort needed to complete the task, a represents the resource's availability, and p represents the average productivity of a typical knowledge worker. Equation 5 solves for d.

$$d = \frac{e}{a} / p$$

Equation 4. Single Point Estimate Using Availability and Productivity

Example

$$d = \frac{10h}{50\%} / 70\%$$

$$d = 20h / 70\% = 26\ hours$$

Equation 5. Single Point Estimate Using Availably and Productivity Example

Strengths

This technique is helpful when the hours given come with high confidence and the work is fairly routine and easy to calculate.

Weaknesses

We have already seen how difficult it is for most people to articulate with any predictability the accuracy of their timelines. Multitasking, or task-switching, is rampant in our fast-paced culture. Since this work ethic is not likely to change, we must use probabilistic math.

Program Evaluation Review Technique (PERT)

PERT—which is recommended by the Project Management Institute (PMI®) and first developed by the U.S. Department of Defense in the 1960s to help the U.S. government manage numerous contractors on its Polaris weapons program—was originally intended to help engineers assess the variances of their duration forecasts. The team who developed the method wanted to know the mean and variance of the length of the activities making up the Polaris program; this would aid them in identifying the critical path. It is important to note that this technique was reverse-engineered from historical data. This point will become critical when discussing its weaknesses. Engineers were asked to estimate optimistic, most likely, and pessimistic durations for assigned work. The team developing this method needed a way to convert a three-point estimate into an equivalent mean and variance and resolved to use a modified beta distribution.

The team was not concerned with the mathematical form of the distribution. As a result, the final formula did not offer any true statistical evaluation of the number set. In Equation 6, μ represents the PERT mean. The mean is calculated using a pessimistic estimate (p), an optimistic (o), and a most likely (ml). The formula is divided by six to account for each of the six variables. To establish the standard deviation (σ), shown in Equation 7, we use the square root of the two extreme values; we will explore a true σ later. Equation 8 solves for μ.

PERT Equation

$$\mu = \frac{p + (4ml) + o}{6}$$

Equation 6. PERT Formula

$$\sigma = \sqrt{\left(\frac{p - o}{6}\right)^2}$$

Equation 7. PERT Standard Deviation Formula

Example

$$\mu = \frac{72 + (4 \times 43) + 15}{6} = \sim 43.17\ hours$$

$$\sigma = \sqrt{3.08} = 1.75$$

Equation 8. PERT and Standard Deviation Example

Strengths

As you can see, the PERT value plots just right of the mean (about one Sigma); this will always happen and is by design (most normal distribution curves realistically plot in this manner). So what does the standard deviation mean to a Project Manager? In this context, it is a measure of the instability of the estimate itself. The larger the deviation, the less confidence we have in the numbers provided. A small deviation denotes some confidence in the estimate—the optimistic and pessimistic values are closer together. Is this a genuine reflection of a standard deviation? No, but we will get to that later!

So why use a simplified, less accurate, version of a formula that is not designed to study probabilities of time estimates? PMs use PERT and the simplified standard deviation calculation as a means to easily and quickly identify the work that needs completing, who can do it, and a reasonably objective assessment of how long that work might take. The variance helps project managers determine how well thought out the estimates are. The higher the variance, the more likely it is that the project manager will ask for more refined estimates from the team. Popularity notwithstanding, the choice to use beta distribution is explained by a mathematician on the original PERT team (Clark, 1962) in the following way:

> The authors have no information concerning distributions of activity times, in particular, it is not suggested that the beta or any other distribution is appropriate. But, the analysis requires some model for the distribution of activity times, the parameters of the distribution being the mode and the extremes. The distribution that first comes to the authors' mind is the beta distribution.[2]

Clark goes on to explain that one of the features of the beta distribution that makes it an attractive choice is that the mathematical manipulation of the distribution is manageable.

Weaknesses

The PERT formula loses any credible application with its use of σ. There are two fundamental issues related to the PERT. First, the endpoints of the beta distribution ($\mathcal{p}, \mathcal{o}$) are, in reality, stochastic. A stochastic quantity is a random variable whose underlying probability distribution changes its shape over time. The flow of traffic is a typical example of a stochastic quantity. The studies we have already reviewed have demonstrated it is problematic for a person to determine the absolute endpoints of a random variable.

The second issue with using PERT stems from the most likely value; the most likely estimate is the modal value of all the possible values of the task duration and not the mean. As resources attempt to assess the most likely, they are drawing on years of experience and are predisposed to provide a mean of their assessment instead of the mode. A mean is the average while the median is the middlemost value in a series of values. The mode is the most recurring.

Because the PERT formula is reverse-engineered from the original historical data, it made the curve artificially fit the circumstances at the time. It was designed to fit the moment and not all moments. Known as curve-fitting, this can be a useful way to predict future events if the underlying statistical processes stay the same. Since all

[2]Beta PERT origins Broadleaf, http://broadleaf.com.au/resource-material/beta-pert-origins/ (accessed September 10, 2016).

projects are different, including from one day to the next, it falls victim to stochastic or random variables. Thus, the PERT method becomes an unpredictable and problematic statistical planning tool.

Adding Resource Availability and Productivity to PERT

PERT is not a statistical measure; it is an estimation. That said, we may improve our PERT analysis by adding our resource's availability and productivity. Equation 9 is an example of this technique. In this instance, e was replaced with the PERT formula. However, we still have a problem with a true standard deviation.

$$d = \frac{e = (p + (4ml) + o)/6}{a}/pr$$

Equation 9. PERT with Availability and Productivity

Statistical Modeling Technique

Having reviewed techniques that only ask for a single point estimate, formulas that have been mistakenly made popular, and techniques that consider availability and productivity, we may now move on to true statistical modeling. We will still use a three-point estimate but rather than applying an artificial variance and standard deviation, we will use complex math, which we can easily replicate in Microsoft Excel.

Models are simplifications of real-world phenomena that we use to hypothesize, explain, analyze, design, and build outcomes. Statistical modeling is a process where we use math and data to construct equations used to predict how models will behave under certain conditions. Models help us make optimized, more informed, lower-cost business decisions by predicting how things behave. When we want to understand the probability of completing a task on a specified date, given an optimistic, most likely, and pessimistic range of numbers, we must determine a measure of the spread of the numbers. To do this, we must calculate a true standard deviation.

Standard Deviation

A deviation simply means how far spread out a set of numbers is from the norm. The calculation is the square root of the variance of a number set. A small deviation indicates the values are tightly grouped around the mean of the data set while a significant deviation indicates the opposite. To determine the standard deviation, we must know the mean and the variance of the number set.

Variance

A variance is the average of the squared differences from a mean, and a mean is the average of a number set. To determine the variance of an optimistic, pessimistic, and most likely set, we must—for each number—subtract the mean, square the result, and then average the squared differences. When we calculate a mean, we must determine if we want to know the population of the data set or a sample. In population and sample statistics, a population refers to the total set of numbers in the set. A sample data set is just that, a sampling of the full set. Because our concern is the population, this paper will only provide equations and examples for the whole data set. Equation 10 shows us how to formulate the population of a number set. Equation 11 is the equation for a population standard deviation of the variance, which we calculate in Equation 10. In the paragraphs that follow, we will breakdown how this equation works and what the different symbols mean.

$$\frac{\sigma^2 = \sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

Equation 10. Population Variance of a Number Set

$\sqrt{\sigma^2}$

Equation 11. Population Standard Deviation of a Variance

Meaning of Equation Symbols for Standard Deviation

Math equations are often written using Greek letters. Lowercase sigma σ, for example, represents the standard deviation of a population data set. The Greek uppercase sigma Σ is used to indicate the sum of all the values in the number set expressed with a subscript notation. The summation, denoted here as $\sum_{i=1}^{N}$, appears with an index and with limits. The summation is read as, "the sum of the N's, and the sum of all the i's." This is the total number of items in the data set. A subscript notation, in this case X sub i, or X_i, tells us that the value we are looking for is the i value. Subscripts allow us to refer to any value in a set or to all of them at once. The Greek letter mu μ represents the mean.

Example

Let us assume the following scenario: A BA has provided a list of all the estimates for all of the activities she is assigned to. The PM wants to know if the data provided has a reasonable level of confidence. The BA has provided nineteen single point estimates. The steps to calculate the standard deviation are:

1. Determine the mean, which is the average of the data set.
2. Find the sum of the squared distances from the mean by subtracting the mean and squaring the result—for each number in the set.
3. Define the variance by computing the mean of the squared results from step two.
4. Identify the square root of the variance. All results are rounded to two decimal places in Table 3.

Table 3. Standard Deviation Equation Scenario Results

Number of values in the set:	19
Sum of values in the set:	311
Mean of the vales in the set:	16.37
Population Variance:	321.39
Population Standard Deviation:	**17.92**

Step 1: Find the Mean of the Data Set

To find the mean of the data set, we solve for μ. As you can see in Equation 12, we total all of the values in the set and divide by how many items are in the set.

$$\mu = \frac{311}{19} = 16.37$$

Equation 12. Standard Deviation Equation Scenario Step 1

Step 2: Calculate the Sum of the Squared Differences from the Mean

As you can see in Equation 13, for each value in the data set, we subtract the mean of the entire set from the value, square the result, and then sum all of the squared results.

$Count\ (n)$	$Set\ (X)$	$Set - Mean\ (x - \mu)$	$Set - Mean^2 (x - \mu)^2$
1	10	10 - 16.37 = -6.37	$(-6.37)^2$ = 40.56
2	23	23 - 16.37 = 6.63	$(6.63)^2$ = 43.98
3	47	47 - 16.37 = 30.63	$(30.63)^2$ = 938.29
4	10	10 - 16.37 = -6.37	$(-6.37)^2$ = 40.56
5	75	75-16.37 = 58.63	$(58.63)^2$ = 3437.66
6	5	5 - 16.37 = -11.37	$(-11.37)^2$ = 129.24
7	7	7 - 16.37 = -9.37	$(-9.37)^2$ = 87.77
8	5	5 - 16.37 = -11.37	$(-11.37)^2$ = 129.24
9	27	27 - 16.37 = 10.63	$(10.63)^2$ = 113.03
10	3	3 - 16.37 = -13.37	$(-13.37)^2$ = 178.71
11	1	1 - 16.37 = -15.37	$(-15.37)^2$ = 236.19
12	13	13 - 16.37 = -3.37	$(-3.37)^2$ = 11.35
13	32	32 - 16.37 = 15.63	$(15.63)^2$ = 244.35
14	18	18 - 16.37 = 1.63	$(1.63)^2$ = 2.66
15	5	5 - 16.37 = -11.37	$(-11.37)^2$ = 129.24
16	3	3 - 16.37 = -13.37	$(-13.37)^2$ = 178.71
17	8	8 - 16.37 = -8.37	$(-8.37)^2$ = 70.03
18	9	9 - 16.37 = -7.37	$(-7.37)^2$ = 54.29
19	10	10 - 16.37 = -6.37	$(-6.37)^2$ = 40.56
$\sum (X) = 311$		$\sum (x - \mu)^2 = 6106.42$	

Equation 13. Standard Deviation Equation Scenario Step 2

Step 3: Calculate the Variance

Calculating the variance depends on whether or not the set represents the total data (population) or partial data (sample). We are only concerned with the population. Equation 14 calculates the variance, while equation 15 identifies the square root of the variance.

$$\sigma^2 = \frac{\sum (X - \mu)^2}{n} = \frac{6106.42}{19} = 321.39$$

Equation 14. Standard Deviation Equation Scenario Step 3

Step 4: Identify the Square Root of the Variance

$$\sigma = \sqrt{\sigma^2} = \sqrt{321.39}$$

$$\sigma = 17.93$$

Equation 15. Standard Deviation Equation Scenario Step 4

Strengths

A practical application of standard deviation in business analysis and project management is the same for PERT except this application of standard deviation is a valid statistical analysis. We still want to know how reliable our estimates are; we are just using the right tool. A project manager may also use standard deviation on a sample of task estimates and use that result to predict how a full set of project estimates will "look." A low standard deviation shows that the estimates are projectable to the larger group.

In another example, the project team has provided all of their estimates. The PM determines that the mean estimate is much longer than he expected. The PM uses a standard deviation to determine if all the other estimates are close to the mean. If the quotes are close to the average task length, then the PM knows that the team has given estimates indicative of their work; their thinking is similar to each other. Conversely, if a few estimates are far from the mean, the PM knows that the resources who supplied the quotes do not agree with that of the team mean. The PM could then ask if those anomalous estimates are representative of the work or an oversight.

Weaknesses

Standard deviations are useful, but they are not without limitations. The driving purpose of this tool is to measure risk. When information is symmetric—the same distribution on either side of the mean—the results are evenly spread and easy to predict. Statistical modeling tells us that when a normal distribution spreads evenly on either side of the mean, the following is true:

- one μ from the mean, contains, 68.27 percent of all potential results,
- two μ from the mean, contains 95.45 percent of all potential results, and
- three μ from the mean, contains 99.73 percent of all potential results.

Figure 4 is an example of a normal distribution curve where the results are evenly spread out from the mean.

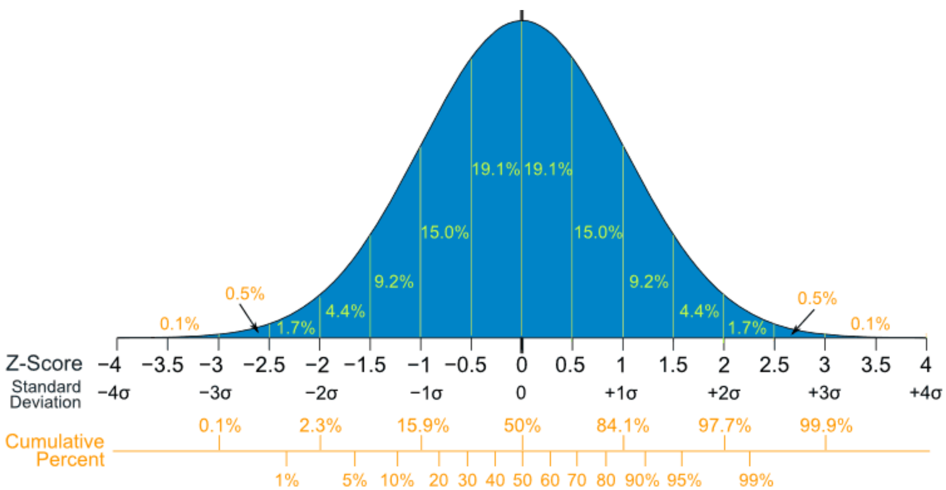

Figure 4. Normal Distribution Curve (Source: https://www.mathsisfun.com)

Many distribution curves are asymmetric, conversely; the results are not spread evenly across the curve. Asymmetric profiles are skewed either positively to the right or negatively to the left. Real-world examples tend to see positively skewed distributions. For these reasons, standard deviation is a measure of volatility, not risk. Applying a probable analysis on an entire project schedule requires a technique that will allow us to account for best, likely, and worst estimates spanning the entire project while including the resource's availability, productivity, and myriad conditions, assumptions, constraints, caveats, and risks, which make project estimations stochastic.

The REPAC Technique

My REPAC Framework (Requirements Elicitation, Planning, Analysis, and Collaboration) uses Bayes' Theorem along with a Cartesian Product to produce the likelihood of a project completing on the estimated date or a date mandated by the client, sponsor, or some other authority. In Relational Algebra, a product is a quantity—in this case a prediction—achieved by multiplying many predictions together from an analogous algebraic operation. Bayes' Theorem accounts for many of the conditions that come with our predictions.

Conditional Bayesian Probability and Bayes' Theorem

Bayesian Probability is a quantity that we assign to represent a state of comprehension or a state of belief. Most elements of a project are highly subjective and conditional. We might provide a quote for an assigned package of work but then specify that, based on our level of comprehension, the assessment is loaded with many considerations such as the number of inter-project dependencies, evolving stakeholder needs, or the time it takes for a stakeholder signatory to return confirmation on a list of stated requirements.

In asking ourselves, "What is the possibility that a stakeholder will agree to come to a requirements workshop?", the response depends on many factors—most of which we cannot control. Project schedules receive estimates that are "conditional" on all of the provisional information we have available to us at the time the estimates are given. As we have learned, our brains are incapable of measuring stochastic values, unpredictable events resulting from random variables. This forces us to take a point-in-time measure. A static estimate, no matter how well thought out it was; it is just that—static. Bayes' Theorem accounts for the randomness of project estimates to help us determine more probable results. Figure 5 describes Bayes' equation. Equation 16 demonstrates the full Bayes' Theorem.

$$P(h|e.b) = \frac{P(h|b) \times P(e|h.b)}{\{P(h|b) \times P(e|h.b)\} + \{P(\sim h|b) \times P(e|\sim h.b)\}}$$

Equation 16. Bayes' Theorem—Full

Probability the assessment is correct = (How expected the evidence is, if the assessment is true) / { (same as above) + (How atypical the assessment is) x (How expected the evidence is, if the assessment isn't true) }

Figure 5. Bayes' Theorem—Translated

Cartesian Product of Sets

The Cartesian product allows us to take multiple sets of project estimates and create one large set (Figure 6), which is used to determine the likelihood that a project will be complete on a projected or mandated day.

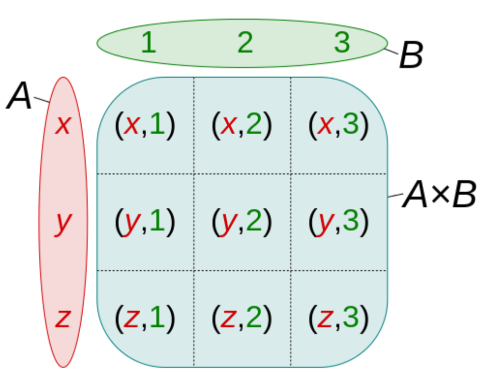

Figure 6. Cartesian Product of Sets (Source: http://www.wikiwand.com)

We read a Cartesian product in the following manner: Columns 1, 2, and 3 in set B join with rows x, y, and z to create A and B, which are each of the nine combined sets. Equation 17 shows us how we build a Cartesian product of the two sets: A and B. In this formula ∈ represents all the units within the set.

$$A \times B = \{(a, b) | a \in A \ and \ b \in B\}$$

Equation 17 Cartesian Product of Two Sets

Finding a Cartesian Product

We have learned how a standard deviation can help us understand how far project estimates are spread out from the mean. Recall, a low standard deviation tells us that most of the numbers are very close to the average and, thus, have a low volatility. A significant standard deviation signifies that the numbers are spread out farther away from the mean—a high volatility. In practice, though, it is more important to know how likely the estimates are.

We will use a Cartesian product of three estimate sets (best, likely, and worst) to build all the possible pairs of sets, which will determine the probability of completing our deliverables for each possible set. We see this type of math all around us and do not even realize it.

To understand how it works, let us imagine a trip to a favorite fast food restaurant. We want to order a meal combo and must decide which options to choose. The following example explains how sets are combined and ordered using a Cartesian product: Burger Combo One (Set A): We may choose from beef, chicken, or pork. Condiments Order Combo (Set B): ketchup, mustard, mayonnaise, or pickles. When joining the two sets together, we see the following results: A x B= {(beef, ketchup), (beef, mustard), (beef, mayonnaise), (beef, pickles), (chicken, ketchup), (chicken, mustard), (chicken, mayonnaise), (chicken, pickles), (pork, ketchup), (pork, mustard), (pork, mayonnaise), (pork, pickles)}. If A and B were reversed, we would simply join the condiments to the meat type—B x A {(ketchup, beef) . . .}.

The REPAC Framework uses three sets of estimates—best, likely, and worst—which are divided by the resource's availability and an average productivity of 72 percent. Joining multiple sets of estimates together allows us to determine the probability of any one set completing within a specified timeframe. Based in sound statistical theory, this distinctive technique allows us to gain a far more accurate picture of our project's timelines. A Cartesian product of three sets combined with a resource's availability and productivity accounts for real-life work situations, shifting deadlines, and changing estimates. Equation 18 demonstrates how the REPAC formula is applied: Let A, B, C represent the best, likely, and worst of an estimate set, and let x, y, z represent a project resource's availability, productivity, and hours in a workday. Therefore, A, B, C are three non-empty project estimate sets.

$$A/(xyz), B/(xyz), C/(xyz)$$

$$A \times B \times C = \{(a, b, c) : a \in A, b \in b, c \in C\}$$

Equation 18. REPAC Product of Sets Technique

Example

Let us examine the following scenario. A business analysis team has submitted best, likely, and worst estimates for all *n* of their tasks. The steps to calculate the probability of completing work within an estimated time using a best, most likely, and worst scenario while considering availability and an average productivity of 72 percent for a complete set of multiple estimates from multiple team members are as follows:

1. For each best, likely, and worst estimate divide by the resource's availability, productivity, and the organization's work hours in a day.
2. Add all of the best, likely, and worst estimates.
3. Using the Cartesian product, find all the ordered pairs in Table 5.

Step 1

Assuming a workday of eight hours and a productivity of 72 percent, for each best, likely, and worst estimate, divide by the resource's availability, productivity, and the organization's work hours in a day. We will keep things simple by using a small amount of sample data. Table 4 represents two business analysis tasks. Equation 19 solves for this table.

Table 4. Business Analysis Tasks to Deliver Example

ID	Task	Worst (hours)	Likely (hours)	Best (hours)	Availability as %
BA-2.14	Plan Workshop: Meta-level Current State	28	16	5.5	25%
BA-2.15	Conduct Workshop: Meta-level Current State	5.5	2.5	2.0	100%

$$A/(xyz), B/(xyz), C/(xyz)$$

$$\frac{\text{BA-2.14}}{\begin{array}{l} A\ 28/(25\%, 72\%, 8) \\ B\ 16/(25\%, 72\%, 8) \\ C\ 5.5/(25\%, 72\%, 8) \end{array}}$$

$$\frac{\text{BA-2.15}}{\begin{array}{l} A\ 5.5/(100\%, 72\%, 8) \\ B\ 2.5/(25\%, 72\%, 8) \\ C\ 2.0/(25\%, 72\%, 8) \end{array}}$$

Equation 19. REPAC Product of Sets Technique Step One

Again, repeat this step for each best, likely, and worst estimate each team member has provided. Table 5 specifies a list of all the best, likely, and worst estimates for each of the tasks within its respective milestone.

Table 5. Business Analysis Milestones to Deliver Example

ID	Milestone	Totals for Best, Likely, Worst		
		Worst (days)	Likely (days)	Best (days)
BA-1	Business Analysis Planning and Monitoring	22.9	43.7	87.8
BA-2	Elicitation and Collaboration	47.9	62.5	100.7
BA-3	Requirements Life Cycle Management	28.8	29.5	51.7
BA-4	Strategy Analysis	11.8	15.3	27
BA-5	Requirements Analysis and Design Definition	46.2	59.4	118.6
BA-6	Solution Evaluation	15.6	19.8	36.8
		2258.00	1692.00	862.00

Step 3

Seen in Equation 20, the Cartesian product has three sets. We find all of the ordered pairs for the sum of each of the best, likely, and worst estimates. Table 5 shows a total of 18 sets. To keep things simple, we will limit our example to the two tasks from Table 4.

$$A \times B \times C = \{(a, b, c): a \in A, b \in b, c \in C\}$$

$$A\{1,2\}, B\{x, y\}, C\{3,4\}$$

$$A\{28,5.5\}, B\{16,2.5\}, C\{5.5,2.0\}$$

$$A \times B = \{(28,16)(28,2.5), (5.5,16)(5.5,2.5)\}$$

$$A \times B \times C = \begin{Bmatrix} (28,16,5.5), (28,16,2.0), (28,2.5,5.5), (28,2.5,2.0), \\ (2,16,5.5), (2,16,2.0)(2,2.5,5.5)(2,2.5,2.0) \end{Bmatrix}$$

Equation 20. REPAC Product of Sets Technique Step 2

We can see that there are eight ordered pairs from just two sets. Neither PERT nor standard deviation alone can help us manage sets of this size.

Step 4

Now that we have a list of all the ordered pairs for all of the best, likely, and worst estimates from all of our project team members, we may now determine the likelihood of each of the estimates of either the team's projected completion date or a mandated one. Table 6 has eighteen number sets, which gives us 729 ordered pairs. Statistically, the probability of any individual outcome is 1/729. First, we order the list from smallest to highest and then use standard deviation and variance for each of the 729 ordered pairs to determine all possible outcomes. For example, if the three smallest possible lengths are 170 days, 171 days, and 172 days, then each of the probabilities is numerically described by the number of desired outcomes divided by the total number of all outcomes.

There is a 1/729 chance the project will be completed in 170 days. There is a 2/729 chance it will be done in 171 days or sooner, and a 3/729 chance it will be done in 172 days or sooner, and so on. This calculation continues for each of the ordered pairs until we reach 100 percent. Table 6 shows an example of probabilities for some of the ordered pairs. This example uses the eighteen sets from Table 5, which calculates into 729 ordered pairs.

Table 6. Probabilities of Ordered Pairs Example

Total Days	Probability
166.32	0.14%
169.79	0.27%
170.49	0.41%
182.99	1.65%
183.68	1.78%
184.38	1.92%
185.07	2.06%
185.76	2.19%
230.21	21.67%
405.90	99.59%
407.64	99.73%
411.11	99.86%
422.92	100.00%

Strengths

The REPAC product of sets technique offers its user a very accurate indication of the likelihood that project deliverables/milestones will complete on the dates specified. The REPAC Technique provides a means for reducing risk with respect to uncertain activities. Uncertain activities are the most difficult to estimate. There is often little data to support a precise estimate. Additionally, there are many factors affecting the estimate such as conflicting schedules, competing projects, and operational duties. Requirements elicitation and analysis is an excellent example wherein stakeholders often have different opinions of needs. The number of iterations needed to reach consensus is often wholly unpredictable.

Weaknesses

Gathering three estimates per task or deliverable can be time consuming and problematic for project teams with strict deadlines. Therein lies the irony. Single point estimates typically offer a 50 percent probability at best. However, where activities are routine, stable, and contain little risk, it may be worthwhile to use more traditional techniques.

Stable activities are those that are well understood and predictable. Stable activity estimating is commonly straightforward. Often, analogous, expert judgment, a parametric model, or published data may be useful. Additionally, project estimates are just that—estimates.

Probability is the measure of the *likeliness* that an event will occur. We are constantly creating hypotheses, making predictions, and then acting on those indications. For example, whether we should bring an umbrella when the forecast is only calling for a 65 percent chance of showers or take the service roads to our destination in the hopes they are free of traffic. Statistical analysis is a powerful tool, but it is just a tool—one of many that BAs and PMs should keep in their toolboxes. The accuracy of any statistical outcome is directly proportional to the quality and quantity of the population size. The more realistic the estimates, the larger the population set (the number of estimated activities), and the more reliable the results will be.

Conclusion

There are many ways to estimate time, and some are more accurate than others. All come with uncertainty. We have a tendency to base our project estimates on the best guess, which makes it difficult to plan for the randomness inherent in these estimates. At best, a single point quote would only give us a 50% likelihood of success. We may increase those odds by looking at a best and worst estimate, plus the resource's general availability and productivity. However, using PERT with these numbers is not a true assessment of possible completion time—it is just a sense of volatility—and not an accurate one at that.

To understand the likelihood of when a project might be complete, we need to use more complex math such as true standard deviation, variance, and a Cartesian product. Even with these tools, timelines are still open to a subjective analysis. Having a strong opinion about an estimate can make it hard to take in new information about it or to consider other options when they are presented. We may reduce our exposure to some of our risk by accounting for the volatility inherent in these quotes by looking at Bayes' Theorem. When dealing with project timelines that are well defined, probabilities can be described by the number of desired outcomes divided by the total number of all outcomes. However, as we have read, projects are not as straightforward as this.

Objectively, a project will finish on the 407th day at 90 percent when given enough chances or iterations. Subjectively, though, the estimates are probably in of themselves—a measure of belief, best guess, or guess-expert opinion. Thus, the use of Bayes' Theorem, commonly known as Bayes' Rule, is an essential ingredient in all project estimations. We argue that the mathematical discovery made by Reverend Thomas Bayes (1701–1761) has given us the power of probabilistic reasoning. It answers for us, "When we encounter new information, how much should it change our confidence in a belief?"

We need not apply a mathematical formula in each of these cases. To achieve Bayesian reasoning we need only remember:

1. How average an estimate is vs. how expected the evidence is
2. If the estimate is true over how atypical the estimate is vs. how expected the evidence is if the estimate is wrong.

Learn More

Learn more about how you can improve productivity, enhance efficiency, and sharpen your competitive edge through training.

[Project Management Fundamentals]

[Schedule and Cost Control]

[PMP Exam Prep Boot Camp]

Visit **www.globalknowledge.com** or call **1-800-COURSES (1-800-268-7737)** to speak with a Global Knowledge training advisor.

About the Author

Perry McLeod is a management consultant, facilitator, and instructor, and author with over fourteen years of experience in business analysis, process reengineering, project management, business modeling, and strategic alignment. Perry delivers industry recognized best practices for some of North America's most successful companies across a number of industries such as banking and finance, agriculture, supply chain, consumer products, software design, insurance, and payment processing. In addition to his many professional accomplishments, Perry was one of the contributors to the IIBA's BABOK® v 2.0.