

# Estimativas de Software - Fundamentos, Técnicas e Modelos

Como usar de forma consistente PF, COCOMOII, Simulação de Monte Carlo e seu bom senso em estimativas de software



## Carlos Eduardo Vazquez

Sócio-fundador da FATTO Consultoria e Sistemas, um dos autores do livro "Análise de Pontos de Função: Medição, Estimativas e Gerenciamento de Projetos de Software", livro com mais de 7.000 exemplares vendidos e atualmente em sua 8ª edição. Pioneiro na aplicação de métricas de software no Brasil possui 20 anos de experiência em TI, notoriamente na aplicação das disciplinas do desenvolvimento e sustentação de sistemas corporativos. Graduado em Processamento de Dados pela PUC-RJ em 1990, já passou com sucesso por quatro vezes pelo processo de certificação de especialista em pontos de função pelo IFPUG – International Function Point Users Group, tendo sido um dos primeiros brasileiros a conquistar essa certificação em 1996. Desde 1993, vem formando profissionais na aplicação da Análise de Pontos de Função, tendo sido professor da UFES, atuado como consultor de grandes projetos de tecnologia em empresas do setor financeiro, bancário e de telecomunicações.

### Neste artigo veremos

Estimar é uma atividade cotidiana, sistematicamente evitada por aqueles responsáveis pela sua execução. Na busca por superar isso, uma série de técnicas e ferramentas surge no cenário do desenvolvimento e manutenção de sistemas. Muitas vezes, no desespero por uma solução imediata, elas são adotadas independentemente de sua adequação ao cenário específico em que serão introduzidas, ou mesmo apenas com um conhecimento superficial quanto ao seu funcionamento. Ferramentas como o COCOMOII, Simulação de Monte Carlo e Pontos de Função não substituem a analista responsável pela estimativa, que enfrenta a confusão entre o que seja uma estimativa técnica, um compromisso pessoal ou uma meta corporativa.

### Qual a finalidade

Nosso objetivo é diferenciar entre esses diferentes atos e como se portar diante de

cada um deles; destacar que simples cuidados podem ajudar a produzir estimativas de muito mais qualidade; apresentar como funciona uma série de ferramentas isoladamente e como integrá-las no estabelecimento de um ambiente propício à melhoria contínua da qualidade das estimativas.

### Quais situações utilizam esses recursos?

Nas diferentes situações em que um analista deve se relacionar com seus clientes no sentido de fornecer a sua expectativa para prazo, custo, esforço ou escopo no desenvolvimento e manutenção de software. Visa ajudar a esse analista a identificar os diferentes tipos de solicitação e evitar que ele caia em armadilhas que o leve a assumir compromissos inexecutáveis. Adicionalmente, é útil também àquele profissional que trabalha na definição de processos de desenvolvimento e seleção de métodos e ferramentas para fins de melhorar o processo de estimativa de sua organização.

### Dificuldades ao Estimar

Estimar é um ato cotidiano na vida de todos e isso não é diferente no desenvolvimento e manutenção de sistemas. Apesar disso, uma série de dificuldades de diferentes naturezas faz com que muitos profissionais

evitem o seu exercício, ou continuamente adiem a comunicação dos seus resultados. Para ilustrar cinco dessas dificuldades, considere um seguinte pedido por uma estimativa: Qual a duração do deslocamento entre o Rio de Janeiro e Niterói?

## Ambigüidade, volatilidade ou falta de clareza

A primeira dificuldade em atender esse pedido é que **(1) os objetivos da estimativa, seu escopo, seus requisitos, não estão claros ou completos**. Por exemplo, qual o meio de transporte para esse deslocamento? Quando se diz Rio de Janeiro, deve-se entender o centro da cidade ou a Barra da Guaratiba (bairro situado a cerca de 60 quilômetros de distância do centro do Rio de Janeiro)? Analogamente, Niterói se refere ao centro da cidade ou a algum outro ponto? O trajeto será feito utilizando um carro ou alguma outra combinação de transportes públicos e privados? Em qual horário?

Veja que essas questões não demandam um grande volume de análise, mas ainda assim muitas vezes elas são desconsideradas, o mesmo acontecendo na prática do desenvolvimento e manutenção de sistemas.

## Garantir medições adequadas e referências válidas

A ambigüidade, volatilidade ou falta de clareza do objeto da estimativa, assim como um domínio de problema não muito bem compreendido, não podem ser empecilhos na realização de uma estimativa, afinal estimar também é relativo a prever face à incerteza.

O conhecimento incompleto não deve ser uma barreira, desde que existam referências nas quais o analista possa se basear. No processo de estimativa, o projeto ou demanda é fracionado em subconjuntos que, individualmente, podem ser designados como uma unidade para execução. Daí surgem outras duas dificuldades: **(2) Garantir que os pacotes de trabalho tenham sido adequadamente medidos;** e **(3) produzir estimativas que estejam consistentes com realizações passadas em outros projetos.**

A construção dessas referências requer dados históricos ou o levantamento dos mesmos – é importante identificar quando não há vontade política para isso, mesmo quando há subsídios técnicos para tal, afinal esse tipo de iniciativa transfere poder de indivíduos para a corporação. Para que esses dados tenham valor efetivo para o processo de estimativa, é necessário que haja um tratamento estatístico dos mesmos, que dificilmente é alcançado na esfera do indivíduo, podendo ser empreendido como uma iniciativa organizacional. Isso porque é tipicamente aí que: (a) Processos são estruturados para estimar e descrever o tamanho do produto de software; (b) as estimativas de custo, esforço, prazo e escopo são relacionados aos valores realizados; (c) os modelos de estimativa utilizados são calibrados às condições locais; e (d) os critérios para normalizar as diferenças entre os projetos e produtos são estabelecidos de tal forma que uma simples extrapolação, não normalizada, de taxa de entrega (Homem-Hora / Ponto de Função) não seja a base para a estimativa.

## Diferenciar estimativa, meta e compromisso

Entre aqueles que pedem e fornecem estimativas, existe muita confusão entre o que seja: a) fornecer uma estimativa, um ato técnico que pondera os riscos de escopo e produtividade; b) estipular uma meta para o atendimento de uma demanda ou projeto por uma equipe, um ato gerencial ou político; c) assumir um compromisso, uma decisão pessoal.

Uma estimativa é uma avaliação do provável resultado quantitativo de uma variável de interesse, é a representação

de uma chance, um número com uma possibilidade de ser realizado, enquanto meta é um objeto a que se dirige algum intento, e compromisso é uma obrigação tácita ou explícita que pode envolver outras pessoas ou ser auto-imposta.

Outra dificuldade ao estimar é que **(4) nem sempre é fácil distinguir entre esses diferentes atos e a importância dessa distinção está na forma como eles são julgados e os fins para os quais os seus produtos são usados**. Uma estimativa não é um orçamento mal realizado, é uma etapa preliminar. Por exemplo, eliminadas as ambigüidades, nivelado o entendimento necessário para estimar a duração do deslocamento entre o Rio de Janeiro e Niterói, identificadas referências válidas, compatíveis com esse entendimento, ainda assim é um absurdo oferecer uma estimativa de uma hora e quinze minutos, pontual (single point estimate) sem o destaque que se trata de uma chance.

## Estimativas diretas versus estimativas paramétricas

As estimativas pontuais são típicas, ainda que não exclusivas das estimativas diretas. Uma estimativa direta é aquela cuja grandeza de interesse (esforço, prazo, custo ou escopo) tem o seu valor estimado de forma direta, sem a utilização de algum outro parâmetro de referência, como no exemplo utilizado anteriormente de 01:15. Um contra-exemplo de uma estimativa direta é como quando se verifica que a distância entre o Rio de Janeiro e Niterói é de 25 quilômetros, considerando uma velocidade média de 25 quilômetros por hora fazendo o percurso todo de carro, e por volta das 15:00, a estimativa seria de 1:00 hora. Esse tipo de estimativa, em que algum outro parâmetro de referência é utilizado para derivar o valor para a grandeza de interesse e é denominada estimativa paramétrica. As estimativas paramétricas são resultados de modelos de estimativas, desde os mais simples como esse exposto, que também produziu uma estimativa pontual, até aqueles que incluem componentes de incerteza.

O aspecto humano é outro fator que deve ser considerado ao receber e elaborar estimativas diretas. Estimar envolve a auto-estima e o reconhecimento profissional. Existem aqueles profissionais que são “orientados ao sucesso”; a estimativa direta produzida por pessoas com esse perfil tende a ser conservadora enquanto aquela produzida por profissionais “viciados em adrenalina” tendem a ser agressivas.

## Modelos determinísticos

Os modelos de estimativa determinísticos não incluem qualquer forma de aleatoriedade ou probabilidade em sua caracterização. Independentemente de sua complexidade, as estimativas são determinadas assim que as suas entradas são definidas. Ao oferecer essas estimativas de 01:15 ou 01:00 para a duração do deslocamento entre o Rio de Janeiro e Niterói, aquele que a recebe pode muito facilmente entender que haja certeza dela estar correta.

A estimativa é obtida de forma direta no primeiro caso (01:15) e a partir de um parâmetro, como a média das velocidades, no segundo caso (01:00). Essa média, por sua vez, é obtida a partir de diferentes viagens passadas onde individualmente verificou-se uma determinada velocidade média no percurso. A **Fórmula 1** descreve como ela foi obtida.

$$\overline{Velocidade} = \frac{\sum_{i=1}^n Velocidade_i}{n}$$

**Fórmula 1.** Cálculo da média das velocidades em n diferentes viagens escolhidas aleatoriamente entre aquelas com as condições similares à que se deseja estimar.

A média foi a medida de tendência de centro utilizada na estimativa de 01:00 para a viagem. Essa medida é afetada por valores extremos que podem diminuir a sua representatividade para fins de estimativas. Por exemplo, se houve um dia nessas viagens com uma sequência de grandes engarrafamentos e, como consequência, a viagem durou três horas, a média terá o seu valor majorado já que dificilmente haja uma viagem cuja duração seja curta o suficiente para haver uma compensação.

Outra medida de tendência menos sensível aos valores extremos é a mediana que, considerando uma lista ordenada das diferentes velocidades verificadas, é o elemento central dessa lista no caso dela conter um número ímpar de elementos, ou a média dos dois elementos centrais caso ela contenha um número par de elementos. Uma terceira medida de tendência é a moda. Ainda considerando a lista citada anteriormente, ela é a velocidade mais comum verificada. Por exemplo, se a velocidade de 15 Km/h foi verificada em três ocasiões e nenhuma outra foi verificada mais vezes, ela é a moda.

Independentemente da medida de tendência, no caso a média, o resultado ilustra a aplicação de um modelo determinístico que, assim como em desenvolvimento e manutenção de software, deveria ser a manifestação de uma probabilidade, e não uma certeza. Como tal, uma estimativa para esses fins deve sempre incluir alguma indicação do quão próximo se espera estar do real, de qual a sua acuidade, por exemplo, mais ou menos 10% ( $\pm 10\%$ ).

## Modelos estocásticos

Os modelos de estimativa estocásticos são aqueles que incluem em sua formulação componentes de incerteza, as suas saídas são verdadeiras estimativas de um sistema real, tendem a ser uma melhor representação da realidade dada a sua inerente natureza aleatória. Em um modelo estocástico, também se considera o grau de dispersão verificado entre as diferentes viagens utilizadas no cálculo da média das velocidades do percurso.

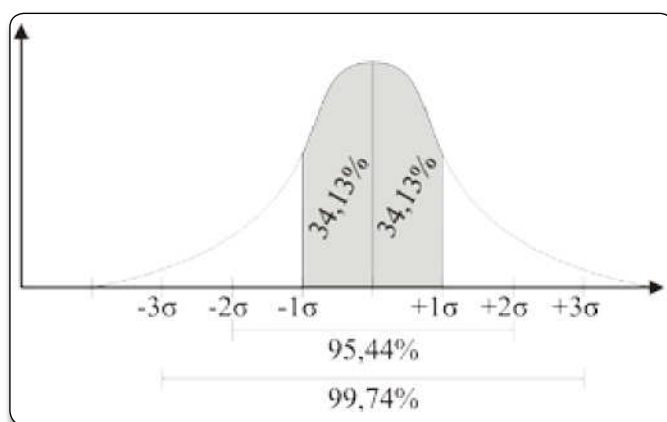
A medida mais usada para descrever esse grau de dispersão é o desvio padrão. Ele é calculado conforme a **Fórmula 2**.

$$s = \sqrt{\frac{\sum_{i=1}^n (Velocidade_i - \overline{Velocidade})^2}{n - 1}}$$

**Fórmula 2.** Cálculo do desvio padrão em n diferentes viagens escolhidas aleatoriamente entre aquelas com as condições similares à que se deseja estimar.

A **Fórmula 2** pode assustar um pouco num primeiro momento, mas para os fins desse texto pode ser resumida como a média dos desvios de cada velocidade individual ( $Velocidade_i$ ) em relação à média das velocidades ( $\overline{Velocidade}$ ). O desvio padrão é fundamental em modelos estocásticos de estimativa e análise de risco, usado como uma unidade e representado pela letra sigma ( $\sigma$ ).

Por exemplo, a soma dos casos com até 1  $\sigma$  de diferença em relação à média em uma distribuição normal representa 68,26% do total dos casos. Em outras palavras, há uma chance de escolher um caso que esteja até 1  $\sigma$  de “distância” da média de 68,26% ou então  $\pm 34,13\%$ . Para 2  $\sigma$ , esse percentual sobe para 95,44%. A **Figura 1** ilustra a distribuição normal e as probabilidades conforme se aumenta a amplitude da faixa de confiança.



**Figura 1.** Distribuição de probabilidades considerando uma distribuição normal onde média, moda e mediana convergem para um mesmo ponto e os extremos se compensam mutuamente.

Considere que o eixo X da **Figura 1** represente as diferentes produtividades verificadas em ordem crescente. Como cada viagem tende a ter uma velocidade única, não é necessária uma precisão nesse nível ao estimar. O mais adequado é que no eixo X haja uma faixa ao invés de uma velocidade individual. Por exemplo, o primeiro elemento dessa série é referente à velocidade entre 5 e 10 Km/h, o segundo elemento, para velocidades a partir de 10 Km/h até 15 Km/h, e assim por diante. Observe que a velocidade não deixa de ser uma unidade de produtividade que relaciona um produto – quilômetros percorridos (Km) – com uma unidade de tempo ou custo – hora (h).

Se o exemplo utilizado fosse o desenvolvimento ou manutenção de software, essa produtividade seria adequadamente representada pelo seu inverso – a taxa de entrega expressa pela quantidade de homens-hora (Hh) por pontos de função (PF), verificada nas diferentes demandas consideradas no estudo.

No eixo Y, está a quantidade de vezes em que o deslocamento em particular teve a velocidade enquadrada na respectiva faixa do eixo X. Aproxima-se bastante do desenvolvimento e manutenção de software onde, ao invés da quantidade de vezes, considera-se a soma das horas apropriadas em cada demanda ou projeto, conforme a respectiva faixa de produtividade.

Qual o normal? Que na maior parte das viagens, haja uma convergência para a média. Haverá casos em que uma viagem



foi feita em muito mais tempo e casos em que ela foi feita em muito menos tempo, mas isso será uma exceção. Daí o nome da distribuição apresentada na **Figura 1**: Normal.

Portanto, se o desvio-padrão fosse de 00:10, a estimativa entre 00:33 (00:43 – 00:10) e 00:53 (00:43 + 00:10) teria 68,26% de chance de estar correta. O desvio-padrão é uma forma de determinar o grau de acuidade em uma estimativa.

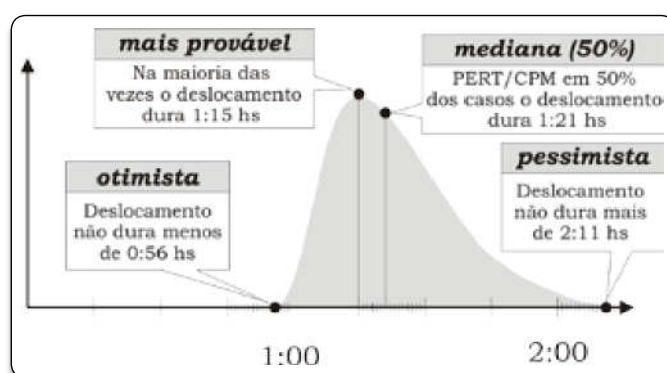
## Outras formas de representar e lidar com a incerteza

O momento em que se estima é implicitamente (e muitas vezes explicitamente) uma indicação de quanta acuidade pode-se esperar. Por exemplo, uma estimativa ao final de um estudo de viabilidade tende a ter uma menor acuidade que outra feita ao final da especificação de requisitos e essa, por sua vez, menor acuidade que uma terceira feita ao final do projeto de alto nível. Esse aumento na convergência entre as estimativas otimistas e pessimistas acontece em função da eliminação tanto de riscos de escopo como de produtividade.

A Associação para o Avanço da Engenharia de Custos Internacional (Association for the Advancement of Cost Engineering International – AACE) recomenda um sistema de classificação de custo específica para a indústria de processos, resumida na **Tabela 1**.

A Associação Americana de Engenheiros de Custo (American Association of Cost Engineers) define três tipos de estimativa de custo: a Estimativa de Ordem de Grandeza ou de Estudo (20 a 25%); Estimativa Preliminar (10 a 15%); e Estimativa Definitiva (5%). Os valores entre parênteses indicam a acuidade esperada nessas estimativas. Em diversos documentos de preparação para certificação como Profissional em Gerência de Projetos (Project Management Professional – PMP) pelo Instituto de Gerência de Projetos (Project Management Institute – PMI), são citados os seguintes tipos de estimativa: Estimativa de Ordem de Grandeza (-25% a +75%); Estimativa de Orçamento (-10% a +25%); e Estimativa Definitiva (-5% a +10%). As faixas entre parênteses também indicam a acuidade da estimativa,

destacando um piso (percentual negativo) que representa uma estimativa otimista e um teto (percentual positivo) que representa uma estimativa pessimista. A Figura 2 ilustra essa dinâmica considerando uma Estimativa de Ordem de Grandeza de 01:15 minutos para o deslocamento entre o Rio de Janeiro e Niterói. Nessa figura, ao invés de uma distribuição normal exposta na Figura 1, utiliza-se uma distribuição beta, que modela eventos cuja materialização é restrita a um intervalo definido por um valor máximo e um mínimo.



**Figura 2.** Distribuição de probabilidades considerando uma estimativa de Ordem de Grandeza de 01:15 (-25% a +75%).

## Usando estimativas para assumir um compromisso

Uma estimativa não é resultado de uma decisão ou resolução, mas estabelecer uma meta ou assumir um compromisso com determinado prazo, esforço, escopo ou custo é uma decisão. Constrói-se uma realidade e trabalha-se para que ela se mantenha. Serem tomadas com base em informação de qualidade, em uma estimativa de qualidade, é um fator que facilita a manutenção dessa realidade planejada.

Por exemplo, um cliente liga para uma central de rádio-taxi pedindo um taxi com hora marcada para Niterói; fornece as informações necessárias e pergunta com que antecedência é

Classe de Estimativa	Nível de Definição do Projeto	Características Secundárias			
		Uso Final	Metodologia	Acuidade Esperada	Esforço de Preparação
	Expresso como % da definição completa	Típico Propósito da Estimativa	Típico Método de Estimativa	Típica variação entre a base e o teto das faixas	Grau de esforço relativo ao menor índice de custo de 1
Classe 5	Zero até 02%	Planejamento do Conceito	Fatoração da capacidade; modelos paramétricos; julgamento; ou analogia	B: -20% a 50% T: +30% a 100%	01 (0.005% do custo do projeto)
Classe 4	01% até 15%	Estudo de Viabilidade	Fatoração do equipamento; ou modelos paramétricos	B: -15% a 30% T: +20% a 50%	02 a 04
Classe 3	10% até 40%	Orçamento, Autorização ou Controle	Custos unitários semi-detalhados com itens de linha de montagem	B: -10% a 20% T: +10% a 30%	03 a 10
Classe 2	30% até 70%	Controle ou Proposta	Custos unitários detalhados com estimados	B: -5% a 15% T: +5% a 20%	04 a 20
Classe 1	50% até definição completa	Verificação da Estimativa	Custos unitários detalhados com reais	B: -3% a 10% T: +3% a 15%	05 a 100

**Tabela 1.** Classes de estimativas para a indústria de processo de acordo com a AACE

mais seguro pedir um taxi para chegar lá. Ela soube de uma história em que outra atendente perdeu o emprego por causa da reclamação de um cliente que perdeu um voo por ter considerado a recomendação de antecedência que ela dera, por isso ela sugere que o cliente marque um horário com uma antecedência de 02:11.

A técnica de estimativa denominada Estimativa de Três Pontos considera não apenas um ponto, mas diferentes pontos que representam a estimativa otimista, mais provável e pessimista. Ela pode ser obtida pela utilização de modelos econômicos que incluam faixas de acuidade pré-estabelecidas conforme o domínio que se deseja estimar (aos moldes daquelas expostas anteriormente) ou solicitando esses pontos para as partes responsáveis pelos diferentes pacotes de trabalho envolvidos. A estimativa otimista considera que tudo sairá tão bem quanto possível e a pessimista que o pior aconteça, porém devem-se desconsiderar eventos muito remotos.

Ao estimar projetos de software, recomenda-se dividir para conquistar, dividir o todo em pacotes de trabalho menores que possam individualmente ser estimados com maior acuidade que o todo. Independentemente da grandeza que se deseja estimar, recomenda-se que cada um desses pacotes de trabalho tenha até 40 horas de esforço. Isso dificulta essa abordagem em momentos preliminares do ciclo de vida de um projeto, o que destaca a importância de haver modelos econômicos para o domínio da engenharia de software.

## PERT/CPM

Na Figura 2 está destacado um ponto (01:21) que representa a mediana entre a estimativa otimista e a pessimista; esse ponto idealmente representa a estimativa cuja possibilidade de subestimar é a mesma de superestimar (50%-50%). Idealmente porque ele foi obtido por uma simplificação definida no modelo PERT/CPM (Program Evaluation and Review Technique/ Critical Path Method) que calcula a média ponderada entre a estimativa otimista, a estimativa pessimista e a estimativa mais provável (ou a moda e cujo peso nessa ponderação é quatro vezes o peso da estimativa otimista e da pessimista), conforme a **Fórmula 3**.

$$Mediana = \frac{Otimista + 4 \times Moda + Pessimista}{6}$$

**Fórmula 3.** Simplificação para obter a mediana (50% de chance) usando o modelo PERT/CPM.

Ao utilizar a fórmula do PERT/CPM, a estimativa para a duração do deslocamento entre o Rio de Janeiro e Niterói é expressa como 01:21 com 50% de chance. Para muitos fins, essa chance pode ser suficiente em tempo de planejamento, esse não era o caso da atendente da central de rádio-taxi!

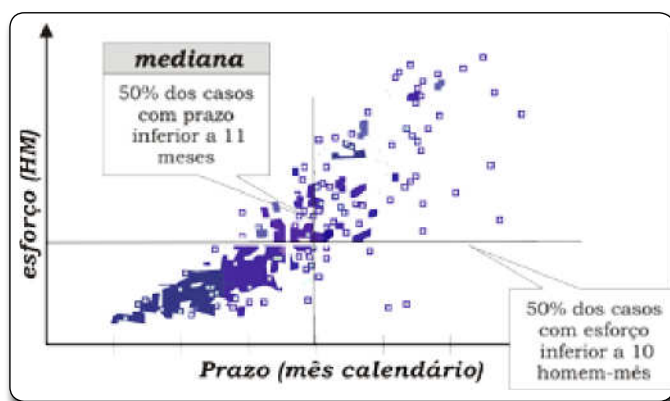
## Simulação de Monte Carlo

Uma terceira técnica que permite obter estimativas pontuais com um nível de confiança, com um percentual de chance nos moldes do exposto na fórmula do PERT/CPM, ou utilizando

uma distribuição normal, é a Simulação de Monte Carlo. Ela também requer como entrada as estimativas otimista e pessimista para a grandeza que se deseja estimar, sejam elas obtidas pelo julgamento de um especialista, pela resolução de um grupo de especialistas ou pela utilização de algum modelo de custos.

O projeto é dividido em partes. Cada parte tem a sua estimativa otimista e pessimista elaborada. A simulação consiste em gerar milhares de valores aleatórios entre esses pontos de cada parte, sumarizar o resultado e avaliar a distribuição da amostra gerada.

Por exemplo, voltando ao contexto do desenvolvimento e manutenção de software e considerando que a grandeza que se deseja estimar seja o prazo conjuntamente com o esforço, a Simulação de Monte Carlo permite fornecer uma estimativa pontual de treze meses com uma chance de 75%, ou então de nove meses com 25% de chance. A Figura 3 ilustra os resultados de uma Simulação de Monte Carlo combinando a avaliação de duas grandezas: esforço e prazo.



**Figura 3.** Resultado da Simulação de Monte Carlo combinando duas grandezas de interesse, esforço e prazo; em destaque o ponto com a mediana de ambos.

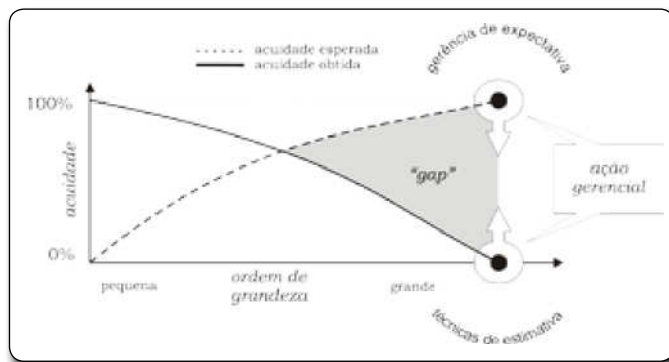
A utilização da Simulação de Monte Carlo não traz resultados relevantes para o processo decisório onde a estimativa esteja num nível de detalhamento muito baixo, com pouco detalhamento. Por exemplo, numa fase bastante preliminar do projeto, não há ainda visibilidade de atividades cujo esforço seja de até 40 horas cada. O seu escopo já está definido, porém ainda não há uma Estrutura Analítica de Projeto (EAP) no nível de granularidade adequado para estimativas diretas de esforço, prazo ou custo. Para suprir a necessidade de estimar também em momentos como esse, os modelos paramétricos foram concebidos.

## A Ordem de Grandeza: Uma quinta dificuldade ao estimar

Com base no que foi exposto, percebe-se uma quinta dificuldade ao estimar: **(5) A ordem de grandeza daquilo que será estimado.** Quanto maior essa ordem de grandeza, maior a dificuldade em produzir uma estimativa de qualidade. Como consequência disso, uma dinâmica perversa se constrói: Quanto maior a ordem de grandeza daquilo que se deseja estimar, maior é a acuidade exigida e mais difícil é obter essa acuidade usando

estimativas diretas. A **Figura 4** ilustra essa dinâmica e o gap que se forma a partir da interseção das duas curvas.

Observe que a acuidade exigida é de 100% num extremo da **Figura 4**. Tecnicamente no desenvolvimento e manutenção de software, isso é inexecutável. Se essa ilustração deixa de ser apenas uma tendência e reflete a realidade em que o analista está inserido, é necessária gerência de expectativas no sentido de diminuir esse gap e foge do escopo deste texto, mas tendo tanta, se não mais importância, que as técnicas aqui apresentadas.



**Figura 4.** Ilustração empírica da relação entre a dificuldade em estimar, a acuidade obtida, e a maior acuidade esperada conforme aumenta a ordem de grandeza do objeto estimado.

Os modelos de estimativa paramétricos dão insumos para o aumento da acuidade da estimativa e a criação de um ambiente propício a isso, principalmente em projetos de maior porte.

## Modelos paramétricos em desenvolvimento e manutenção de software

Iniciativas de introdução de modelos de estimativa paramétrica no desenvolvimento e manutenção de sistemas visam maior assertividade e mais subsídios para uma negociação informada, onde haja trocas e concessões entre as diferentes variáveis do projeto: Se um prazo 20% menor é necessário, quanto do escopo pode ser adiado para outra versão? Por que o custo apresentado é de R\$ 120.000,00 e não metade? O que é necessário para que essa seja a estimativa de custo para o projeto? Ao empreender uma iniciativa como essa, respostas são seguidas de uma explicação com base em dados históricos e análise de probabilidades.

Para isso, a sistematização do planejamento do escopo (quilômetros percorridos em nosso exemplo inicial) é um passo necessário e seu início se dá pela introdução de uma unidade de tamanho para o escopo, um fator primário de custo. A partir dessa unidade é que será possível estabelecer as relações de esforço, custo e prazo com o escopo, viabilizando o que está sendo denominado neste documento como decisão informada.

## Relação entre escopo e esforço

O fator primário de custo, essa unidade de escopo, será a base para a estimativa de esforço e, associado a outros fatores secundários que aumentem ou diminuam a produtividade observada, são as entradas para modelos de estimativa paramétrica. A maior parte deles utiliza a **Fórmula 4** como base

de sua elaboração.

$$M_{\text{min } t} = A \times \text{Tamanho}^B$$

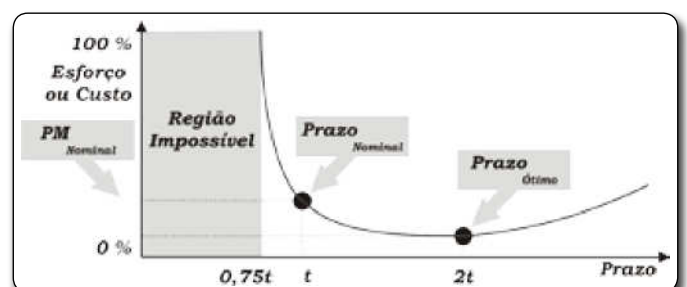
**Fórmula 4.** Fórmula base para estimativa de esforço.

O Tamanho é o fator de custo primário citado anteriormente, o esforço estimado nessa fórmula está representado pela sigla PMnominal. Ele tem a denominação PM em função da unidade tradicionalmente usada nesse contexto ser o Homem-Mês (Man-Month ou o termo politicamente correto e mais atual Person-Month). O nominal que o qualifica deve-se à consideração de um esforço que não incluía a sobrecarga advinda da compressão ou distensão de cronograma, da solicitação de um prazo mais curto ou mais longo do que aquele onde se equilibra a relação entre prazo e esforço.

Por exemplo, considera-se um desenvolvimento cujo esforço é originalmente estimado em 06 pessoas-mês e o prazo em 03 meses-calendário sendo empreendido com meta de prazo inferior. Ao apresentar essa estimativa, parece razoável que haja uma relação linear entre o esforço, o prazo, e a quantidade de recursos mobilizados. Com base nessa lógica podemos assumir que o mesmo trabalho descrito anteriormente será executado em um prazo de 01 mês-calendário por uma equipe de 18 pessoas.

Frederick Philips Brooks denominou essa lógica de “O Mítico Homem-Mês” (mesmo nome dado ao livro clássico The Mythical Man-Month: Essays on Software Engineering publicado em 1975, mas bastante atual em diversos aspectos). A partir de determinado ponto, na medida em que se colocam mais pessoas em um projeto, isso aumenta o custo em um ritmo superior ao que se aumenta a produção, em outras palavras, o prazo não se reduz proporcionalmente a essa mobilização.

A pesquisa de Putnam Norden identificou que, para projetos onde comunicação e aprendizado são muito necessários (como projetos de software) a curva que representa a relação entre o esforço e o tempo segue a distribuição de Rayleigh. Putnam confirmou que esta curva se aplica aos projetos de software. A curva resultante ficou conhecida como Curva Putnam Norden Rayleigh ou Curva de Staffing PNR, ilustrada na **Figura 5**.



**Figura 5.** Curva Putnam Norden Rayleigh – PNR.

Os modelos de estimativa como o descrito na **Fórmula 4** produzem estimativas de esforço (PMnominal) referentes a um determinado prazo (t) representando um cenário onde as prioridades entre esforço e prazo estão equilibrados. O esforço



ótimo (e que compromete o prazo) é aquele verificado no dobro do prazo resultante do modelo (2t). A **Figura 5** também ilustra que um prazo inferior a 75% do prazo resultante do modelo (0,75t) não é possível de ser alcançado independentemente de quantos sejam os recursos mobilizados para a empreitada.

### Normalizando o esforço entre diferentes projetos

Não existe um padrão que estabeleça quantos Homens-Hora (Hh) equivalham a uma Pessoa-Mês e um cuidado ao analisar dados de produtividade, utilizar ou calibrar modelos de estimativas é verificar se todas as referências estão normalizadas nesse sentido. Por exemplo, um projeto teve o esforço apurado em 20 PM e 160 homens-hora foi considerado como o equivalente a uma pessoa-mês. Para fins de planejamento, trabalha-se com 152 homens-hora. Portanto, para fins de análise dos dados será considerado o esforço de 21,05 PM. A **Fórmula 5** ilustra essa normalização.

$$2,6_{PM} = \frac{(160_{Hh/M} \times 0_{PM})}{152_{Hh/M}}$$

**Fórmula 5.** Fórmula para normalizar o esforço em diferentes projetos onde a quantidade de Hh/PM não é uniforme.

### Normalizando o prazo entre diferentes projetos

De maneira análoga ao esforço, quando se analisam dados históricos, utilizam-se ou calibram-se modelos de estimativas, é importante conhecer o prazo nominal, que desconsidera a compressão ou distensão de cronograma. Por exemplo, se um projeto durou 12 meses-calendário e houve uma compressão de cronograma de 25% em relação ao nominal, deve-se considerar que este prazo nominal seja 16 meses-calendário. A **Fórmula 6** ilustra a normalização.

$$6_{N \text{ min } t} = \frac{2_{Medido}}{3\%}$$

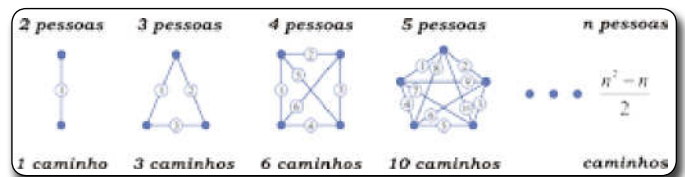
ou

$$Duração_{N \text{ min } t} = \frac{Duração_{Medido}}{SCED\%}$$

**Fórmula 6.** Fórmula para obter o prazo nominal a partir do prazo efetivo e a compressão de cronograma nele verificada

### Capturando a produtividade

As constantes A e B da **Fórmula 4** são os fatores secundários de custo e buscam capturar a produtividade pela ponderação dos efeitos multiplicativos no esforço conforme o escopo aumenta (A) e pelas economias e deseconomias de escala que têm uma natureza exponencial (B). Essas últimas causadas principalmente: a) pelo aumento no esforço em função da sincronização entre as atividades realizadas concomitante-mente; e b) pelo aumento da quantidade de caminhos de comunicação entre os envolvidos na realização dessas atividades. A **Figura 6** ilustra o aumento da complexidade das comunicações na medida em que aumenta a quantidade de pessoas envolvidas.



**Figura 6.** Aumento exponencial da complexidade na medida em que se aumenta a quantidade de pessoas interagindo e trocando informações.

A solução encontrada para lidar com esse tipo de complexidade é a introdução de agentes intermediando, coordenando o trabalho entre essas pessoas, e que não estejam diretamente ligados à produção – o gerente. Ele não produz, contudo, viabiliza a consecução de empreendimentos impossíveis sem esse tipo de ação.

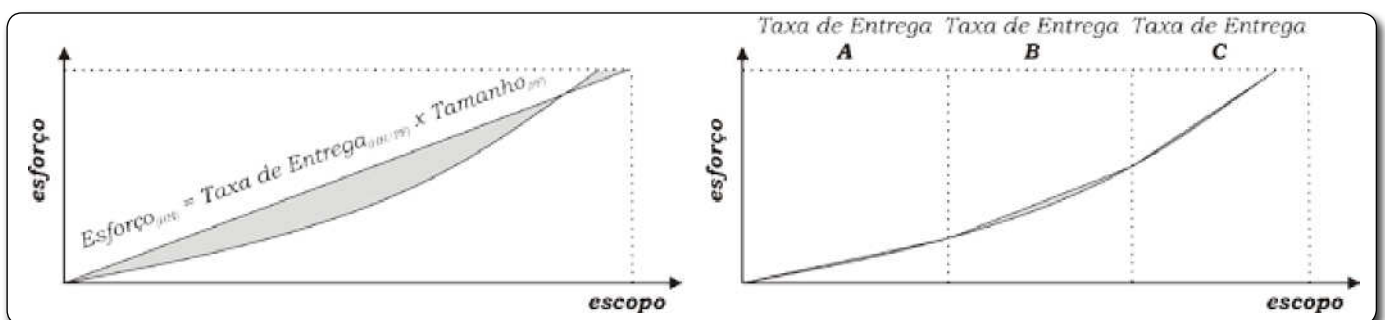
### Modelos baseados em simples relações usando a APF

Os modelos de estimativa baseados na análise de pontos de função (**Fórmula 7**) são uma simplificação onde os fatores de escala, que têm um efeito não linear na produtividade, são desconsiderados e a taxa de entrega busca capturar os efeitos de A e B.

$$Esforço_{(H)} = Taxa \text{ de Entrega}_{(H / P)} \times Tamanho_{(P)}$$

**Fórmula 7.** Estimativa de esforço a partir do escopo estimado/medido em pontos de função.

Se considerarmos a aplicação dessa simplificação em um contexto onde o escopo dos projetos/demandas é relativamente uniforme, por exemplo, estejam entre 200 e 500 PF, essa



**Figura 7.** Adequação de um modelo linear de estimativa de esforço a partir do escopo, mesmo quando a realidade modelada tem um comportamento não-linear usando diferentes categorias de tamanho.

simplificação é plenamente adequada para fins práticos de estimativa, como representado na **Figura 7**.

Portanto, ao usar uma taxa de entrega para estimar o esforço, é importante que ela seja adequada à ordem de grandeza do projeto que se deseja estimar. Adicionalmente, os pontos de função, hoje, buscam representar uma das dimensões do tamanho de uma aplicação ou projeto – a funcional. Essa, porém, não é a única a ser considerada.

### O fator de ajuste e a captura dos aspectos não funcionais

Quando a análise de pontos de função (APF) foi concebida por Allan J. Albrecht entre 1974 e 1979, sua intenção foi medir a produtividade no desenvolvimento de software e, para isso, a definição de uma unidade de produto foi fundamental. O artigo *Measuring Application Development Productivity* define uma técnica cujo tamanho resultante inclui tanto aspectos funcionais, afins às práticas e procedimentos do usuário, quanto aspectos técnicos e de qualidade. Aos primeiros, dá-se o nome de pontos de função não ajustados (Unadjusted Function Point Count) e, aos últimos, fator de ajuste (Value Adjustment Factor ou VAF); a combinação dos dois resulta nos pontos de função.

Naquela época, o UFPC era obtido com base na contagem das funções conforme o seu tipo – entradas, saídas, consultas e arquivos mestres (esses arquivos podem ser vistos como os requisitos de armazenamento da aplicação em contraste com os arquivos de movimento que são a mídia pela qual as transações chegam para processamento em sistemas batch). Cada tipo tinha uma quantidade de pontos de função correspondente, respectivamente – 04, 05, 04 e 10. O VAF era obtido pela ponderação do nível de influência de 10 características gerais de sistema no projeto ou aplicação, sendo medidas em uma escala de zero, indicando nenhuma influência, a cinco, indicando uma característica essencial. Pela totalização dos níveis de influência atribuídos a cada uma dessas características, determinava-se o VAF numa faixa entre 0,75 (causando uma redução em relação ao UFPC de 25%) e 1,25 (aumentando em 25%). Não havia orientação para proceder a essa classificação dos níveis de influência.

Em 1984, a técnica ganha a sua forma atual onde, além dos quatro tipos de função originais, também são contabilizadas as necessidades por dados externos à aplicação, os arquivos de interface externa; cada tipo de função passa a ter sua complexidade funcional avaliada como baixa, média, ou alta; e ambos associados – o tipo de função e a complexidade funcional – determinam a contribuição de cada funcionalidade aos UFPC.

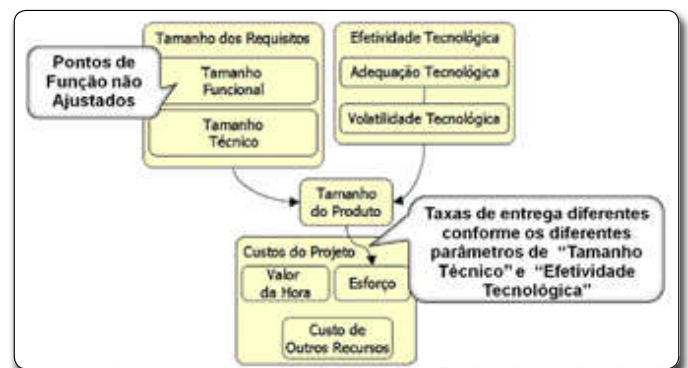
Apesar de se manter essencialmente a mesma desde então, nesses 25 anos tem evoluído no sentido de ser uma métrica funcional que pondera apenas as práticas e procedimentos do usuário. Isso não implica que aquelas características ponderadas pelo VAF (e outras mais) não sejam importantes para o processo de estimativa.

As principais críticas à determinação do VAF e os motivos pelos quais a comunidade de usuários de pontos de função deixa de utilizá-la para ponderar o tamanho técnico do produto são: a) todas as 14 características têm o mesmo espectro de impacto de 5% no resultado da medição e, conseqüentemente,

na estimativa; b) o peso relativo de cada característica é essencialmente estático e parte da definição da técnica não sendo passível de calibração, por exemplo, não é possível considerar que a complexidade do processamento tenha um impacto entre 0,73 (reduzindo a estimativa em 27%) e 1,74 (causando um aumento da mesma em 74%); e c) as orientações em sua determinação refletem um cenário obsoleto e descolado da realidade atual, por exemplo, um aplicativo onde mais de 30% das transações sejam interativas contabiliza o maior nível de influência ao avaliar a característica entrada de dados on-line. A subjetividade, as diferentes possíveis interpretações na determinação dos níveis de influência com base nas orientações fornecidas, não está nesta lista. Isso porque a taxa de entrega utilizada na elaboração da estimativa de esforço numa mesma organização com uma interpretação local, comum, estabelecida, convencionada, cumpre o papel de calibrar o modelo nesse particular.

### O papel das categorias

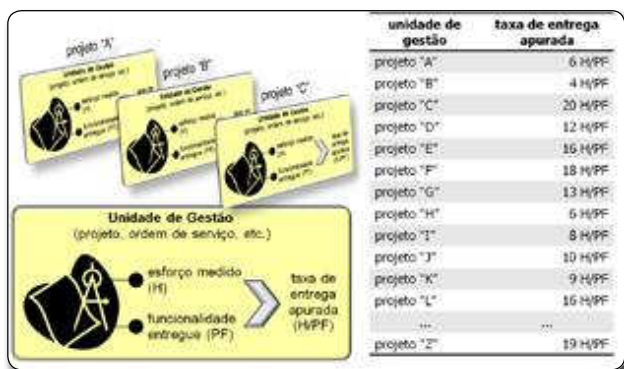
A definição de diferentes categorias de projetos, cada uma com uma taxa de entrega específica, conforme os diferentes contextos representativos dos requisitos de qualidade e técnicos, é uma alternativa para capturar o impacto desses aspectos não funcionais e as economias ou deseconomias de escala usando o modelo com simples relações entre pontos de função e esforço. A **Figura 8** ilustra uma estratégia para isso.



**Figura 8.** Uma estratégia para representar os aspectos não funcionais para fins de estimativas de software.

A prática observada no mercado é a definição dessas categorias apenas com base na linguagem de programação ou plataforma. Contudo, nem sempre essas duas dimensões são suficientes para capturar outros requisitos como, por exemplo: a complexidade do produto; o processo de desenvolvimento ou gestão utilizado; o grau de confiabilidade exigido; entre outros. Para tanto, seria necessário empreender um estudo que buscasse a determinação dessas categorias com base nas diferentes produtividades verificadas e na identificação dos aspectos comuns que levaram projetos diferentes a apresentar o mesmo nível de produtividade. A **Figura 9** ilustra a definição de unidades de gestão na qual atributos são coletados e tabulados visando identificar as diferentes categorias de produtividade verificadas, e os critérios para que um projeto ou demanda seja enquadrado numa categoria.

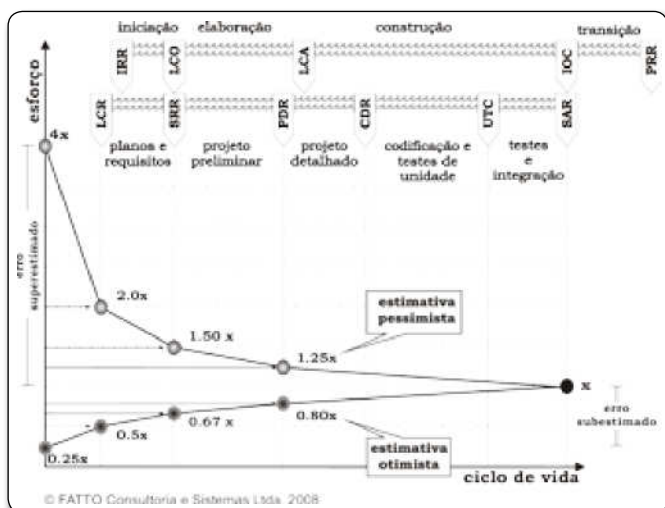




**Figura 9.** Coleta e apuração das taxas de entrega em diferentes projetos em busca das diferentes categorias de produtividade e critérios para o enquadramento de um projeto nessa categoria.

## COCOMOII

Uma alternativa ao empreendimento de um trabalho de definição de categorias, ou pelo menos uma plataforma para isso, é a utilização de modelos pré-definidos de estimativa. O trabalho mais abrangente sobre engenharia de software nessa perspectiva econômica é a definição do modelo denominado Constructive Cost Model (COCOMO), coordenado pelo Professor Barry Boehm na Universidade do Sul da Califórnia (USC), e que originou o livro *Software Engineering Economics* publicado em 1981. O trabalho relativo a essa iniciativa não parou e hoje o modelo encontra-se em sua versão COCOMOII publicada em 2000, assim como houve uma série de suplementos também publicados desde então. Ele é um modelo aberto ao público e bem definido, com uma diversidade de parâmetros adequados aos níveis de informação disponíveis em diversos momentos do ciclo de vida, podendo ser calibrado às condições locais e estruturado conforme os diferentes segmentos do mercado de software.



**Figura 10.** Uniformização dos marcos entre os diferentes processos de desenvolvimento; explicitação das fases; e demonstração da acuidade na estimativa conforme a fase do ciclo de vida de acordo com o COCOMOII

O primeiro valor do COCOMOII está na definição de suas premissas. Elas estruturam de maneira unificada as principais correntes metodológicas na engenharia de software. Essa estruturação consiste de:

(a) Marcos comuns, onde determinados artefatos são entregues; (b) Fases numa perspectiva externa à função de desenvolvimento para fins de planejamento e controle; (c) Atividades cujo escopo, esforço, custo e prazo serão considerados na aplicação e geração de indicadores de produtividade; e (d) Grau de acuidade, a amplitude da incerteza conforme a fase em que a estimativa será realizada.

O modelo não se limita a isso, nem tão pouco esse é o seu objetivo final. Porém, ao fazer isso, constrói todo um arcabouço, toda uma fundação, que facilita a prática de estimativa independentemente da aplicação do modelo propriamente dito. A **Figura 10** ilustra esse arcabouço.

## Conclusão

Este artigo procurou definir o que é uma estimativa e diferenciá-la de outras ações com as quais ela se confunde, apresentar uma breve introdução sobre uma série de ferramentas e técnicas para sua realização, e desfazer alguns mal entendidos sobre a dinâmica do uso da análise de pontos de função em estimativas de software, especialmente o papel do valor do fator de ajuste.

Em um espaço como este seria impossível esgotar o assunto em profundidade. Em próximos artigos vamos especificamente tratar de cada uma dessas técnicas sem perder o foco em como integrá-las com outras técnicas, ferramentas e práticas que visam melhorar a qualidade das estimativas de software, começando pelo COCOMOII. ●

## Referências

- L.H. Putnam, "A General Empirical Solution to the Macro Software Sizing and Estimation problem," IEEE Transactions of SW Engineering, 1978.
- Boehm, Barry et al. "Software Cost Estimation With COCOMO II," Prentice Hall, 2000.
- Robert E. Park, "A Manager's Checklist for Validating Software Cost and Schedule Estimates," CMU/SEI-95-SR-004, 1995.
- Brooks, Frederick P., Jr., "The mythical man-month: essays on software engineering – Anniversary ed.," Addison Wesley Longman, Inc., 1995.
- Edward Yourdon, "Death March: The Complete Software Developer's Guide to Surviving Mission Impossible Projects," Prentice Hall, 1997.
- IFPUG – International Function Point Users Group, "Framework for Functional Sizing," 2003.
- Allan J. Albrecht, "Measuring Application Development Productivity," 1979.
- AACE International, "Recommended Practice No. 18R-97, Cost Estimate Classification System – as applied in engineering, procurement, and construction for the process industries," 2005.
- Thomas A. Willian, "Estatística Aplicada: à Administração e Economia," 2ª Edição, Cengage Learning, 2007.
- Daniel D. Galorath e Michael W. Evans, "Software Sizing, Estimation, and Risk Management," Auerbach Publications, 2006.
- Steve McConnell, "Software Estimation: Demystifying the Black Art," Microsoft Press, 2006.
- Carlos Vazquez, Guilherme Simões, Renato Albert, "Análise de Pontos de Função: Medição, Estimativas e Gerenciamento de Projetos de Software," Editora Érica, 2008, 8ª Edição.

## Dê seu feedback sobre esta edição!

A Engenharia de Software Magazine tem que ser feita ao seu gosto. Para isso, precisamos saber o que você, leitor, acha da revista! Dê seu voto sobre este artigo, através do link:

[www.devmedia.com.br/esmag/feedback](http://www.devmedia.com.br/esmag/feedback)

