

COPE: Correspondência Conceitual entre o modelo PSM e a Estatística

Rafael V. M. Santos^{1†}, Marcelo S. Oliveira², Eric B. Ferreira³, Sabrina B. Vargas⁴, Eros E. Moura¹

¹ Instituto Federal do Espírito Santo (IFES)

² Professor Associado IV, Departamento de Ciências Exatas, Universidade Federal de Lavras (UFLA)

³ Professor Adjunto III, Instituto de Ciências Exatas (ICEx), Universidade Federal de Alfenas (Unifal-MG)

⁴ Universidade do Espírito Santo (UNES)

Resumo: Uma correspondência entre a Estatística e conceitos do Modelo de Medição PSM com a finalidade de garantir a organização estatística dos dados, possibilitando a realização de delineamentos experimentais para a identificação de melhorias a serem propostas a Processos de Desenvolvimento de Software.

Palavras-chave: Modelo PSM; Estatística; Processo de Software.

Abstract: A correspondence between Statistics and the measurement model concepts PSM with the purpose of statistical organization and allowing the design of the experiments for detecting improvements to be made in the software development process.

Keywords: PSM Model; Statistics; Software Process.

Introdução

Este artigo descreve a relevância, a estratégia de desenvolvimento e resultados obtidos na correspondência entre conceitos estatísticos e conceitos do PSM (COPE). Essa correspondência foi concebida como sendo parte de um Processo de Medição de Software, baseado na filosofia Estatística do Seis Sigma, processo este que tem por objetivo auxiliar as empresas no desenvolvimento de seus projetos de software.

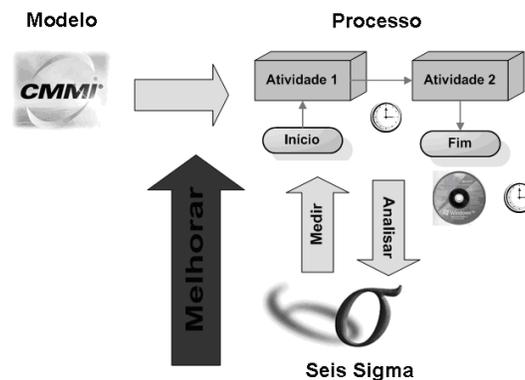


Figura 1: Contexto do Trabalho

[†] Autor correspondente: ravarmes@gmail.com.

O desenvolvimento da COPE é parte integrante de um Processo de Medição de Software, baseado na Filosofia Estatística do Seis Sigma. O Processo de Medição citado anteriormente foi desenvolvido em dissertação de mestrado, no Departamento de Ciências Exatas (DEX) da Universidade Federal de Lavras (UFLA) (SANTOS, 2007).

A Figura 1 introduz o contexto no qual este trabalho está aplicado. Pode-se observar pela mesma que um processo de desenvolvimento de software (Processo) é proposto levando-se em consideração um modelo de processo de desenvolvimento de software (CMMI). O processo de medição de software (representado pelo símbolo do Seis Sigma) visa à medição (Seta Medir), representada pelos relógios em cada uma das atividades do processo, e análise (Seta Analisar) dos dados. O CD representa o produto final, ou seja, o software.

O processo de medição, o qual é representado pelo símbolo do Seis Sigma, dentre outras funcionalidades, provê melhoria no processo de desenvolvimento de software, conforme apresentado pela seta **Melhorar**, também mostrada pela Figura 1. Para que esta melhoria possa ser aplicada no processo de desenvolvimento de software, o Seis Sigma instrui para uma coleta de dados, bem como a análise destes dados por meio da utilização de delineamentos experimentais (DOE – “Design of Experiments”).

Neste contexto temos a COPE. Isso, pois, para que os delineamentos experimentais possam ser realizados, os dados precisam estar organizados de forma a permitir esta análise experimental proposta. Segundo o processo de medição citado, estes dados devem ser armazenados na ferramenta PSM Insight (2013). A COPE tem o principal propósito de organizar estatisticamente os dados coletados do processo de desenvolvimento a fim de possibilitar a realização de DOE's, e consequentemente a melhoria do processo.

Para um melhor entendimento deste artigo as seções a seguir visam introduzir alguns conceitos correspondentes ao ponto de partida da correspondência (COPE) aqui apresentado.

CMMI

De acordo com Paulk et al. (1995) para melhorar a capacitação da indústria de software dos Estados Unidos, foi fundado o SEI, “Software Engineering Institute” (Instituto de Engenharia de Software), junto com a “Carnegie-Mellon University”. O resultado dessa iniciativa materializou-se no desenvolvimento do CMM – modelo de melhoria de processo, baseado nos princípios de controle estatístico da qualidade, estabelecidos inicialmente por Shewhart, desenvolvidos e demonstrados por Deming e Juran, e no conceito de maturidade da gestão de qualidade, estabelecido por Crosby sendo adaptado para os processos de software por Humphrey (1988) e outros.

O CMMI é uma evolução do CMM e procura estabelecer um modelo único para o processo de melhoria corporativo, integrando diferentes modelos e disciplinas.

Paulk et. al. (1995) propõe que a melhoria dos produtos de software seja obtida focando **prioritariamente a melhoria dos processos de software**. Segundo Paulk et. al. (1995) a premissa fundamental da gestão de processo de software é que a qualidade de um produto de software é grandemente determinada pela qualidade do processo utilizado para desenvolvê-lo e mantê-lo.

Seis Sigma

Para Snee (2000) o Seis Sigma é um sistema que liga ideias, tendências e ferramentas, o qual o foco no cliente torna-se a prioridade principal. As melhorias Seis Sigma são definidas pelo seu impacto sobre a satisfação e valores dos clientes.

Os conceitos fundamentais do Seis Sigma consideram o fato de que a variação dos produtos e processos deve ser conhecida por ser um fator que afeta tempos de fabricação, custos de produto e processo, qualidade do produto e, finalmente, a satisfação do cliente. A etapa crucial do Seis Sigma consiste na definição e medição da variação dos processos com o objetivo de descobrir suas causas, desenvolvendo meios operacionais eficientes para controlar e reduzir esta variação (USEVICIUS, 2004).

Medição de Software

O uso de medições nos ajuda a entender e a interagir com o mundo para, então, poder melhorá-lo. Muitas medições são propostas e aplicadas em casos práticos a fim de alcançar os seguintes objetivos: i) melhorar o entendimento sobre o processo, produto, recursos e ambiente de desenvolvimento e, assim, estabelecer bases para comparação entre medições; ii) avaliar o andamento do projeto comparando com dados planejados; iii) fazer previsões sobre o futuro andamento do projeto baseado em comportamentos passados; iv) promover melhorias identificando falhas, ineficiências e outras oportunidades para melhorar a qualidade do produto e o desempenho do processo (PARK et al., 1996).

Porém, ao contrário do que possa parecer, definir, coletar e analisar um conjunto de medições não é uma tarefa trivial. Na realidade, esta é uma tarefa custosa que demanda grande conhecimento para evitar que o seu uso não aumente ainda mais os problemas enfrentados durante o desenvolvimento. De acordo com Park (1996), este problema ocorre devido ao grande número de atributos que podem ser medidos durante um projeto. Uma escolha incorreta das medições pode levar, além do próprio aumento desnecessário de esforço, a uma visão distorcida do processo o que dificulta a sua análise e, muitas vezes, termina por orientar decisões equivocadas.

Neste contexto se destaca o modelo de medição de software PSM e a abordagem Seis Sigma que se baseia na convicção de que para uma organização medir de forma eficiente, é necessário, primeiro, especificar objetivos a serem alcançados, relacionar estes objetivos com dados reais obtidos por intermédio de medições e, finalmente, prover um processo de medição para a interpretação destes dados de acordo com os objetivos propostos.

O artigo considera o CMMI como modelo de processo de desenvolvimento de software, bem com, o modelo PSM (principalmente a ferramenta PSM Insight (2013) e a filosofia Estatística do Seis Sigma, como influenciadores diretos na proposta do processo de medição citado na seção contexto. Considerando a grande importância desses influenciadores diretos, o artigo detalha a COPE. Proposto na finalidade de organizar a armazenagem dos dados coletados durante o desenvolvimento do software, de forma que os mesmos ofereçam a possibilidade de obtenção de informações importantes para as tomadas de decisões.

Nas seções a seguir serão definidos os principais conceitos relacionados a um processo de medição de software, utilizados tanto no modelo PSM, como na Estatística. Os quais tiveram de ser entendidos antes da realização da COPE.

Definição de Conceitos do PSM

Segundo McGarry et al. (2002) o modelo de informação do PSM é uma estrutura para a definição das medidas a serem utilizadas em um projeto. O Modelo define os seguintes conceitos:

- Atributo: propriedade relevante do ponto de vista das necessidades de informação.
- Estrutura: nível de organização das medidas.
- Indicador: estimativa ou avaliação que provê uma base para a tomada de decisão.

Definição de Conceitos Estatísticos

Segundo Stell e Torrie (1980), o delineamento de experimentos (e sua análise de variância associada), introduzida por Ronald A. Fisher está essencialmente voltada a um processo de estudo de soma de quadrados e dos componentes associados a conhecidas fontes de variação, podendo ser usada em todos os campos de pesquisa onde os dados são mensurados quantitativamente.

No âmbito da engenharia de software os experimentos ajudam na identificação de causas comuns de variação do processo, ou seja, causas inerentes às variações do processo. Estas causas não podem ser identificadas por um gerente de projetos. Por meio desta técnica, alterações a serem realizadas no processo de software podem ser propostas com vista à melhoria do processo.

A seguir serão conceituados os principais termos frequentemente empregados no planejamento e na análise de variância.

- Experimento: é o processo que permite a coleta de observações sob condições previamente controladas.
- Dados Experimentais: são as observações obtidas dos experimentos. Os dados experimentais refletem a influência de diferentes variáveis, além do fator em estudo. Estas variáveis são devidas ao ambiente experimental e podem ser controláveis ou não.
- Fator: é a variável cujo efeito se deseja conhecer e avaliar no experimento.
- Resposta ou Variável Resposta: é a variável a ser medida ou avaliada no experimento, gerando os dados.
- Tratamento: o termo tratamento é utilizado para caracterizar os valores (tipos ou níveis) que um fator pode assumir em um determinado experimento.

Estratégia de Desenvolvimento

Primeiramente a visão Estatística voltou-se ao CMMI. Por meio dessa primeira compreensão Estatística do CMMI, mostrado resumidamente pela seção 3.1, observou-se a necessidade de desenvolvimento da COPE, que é o assunto principal desse artigo, o qual é apresentado detalhadamente por uma seção a seguir.

A Correspondência entre CMMI vs Estatística

Uma vez que este artigo está focado em apresentar a COPE, é importante entender a aderência entre o Seis Sigma e as áreas que tratam de medições no CMMI.

Segundo Oliveira (2006) podemos apresentar em alto nível a ligação entre os níveis do CMMI e a Medição por meio da Figura 2. Nesta figura podemos verificar a importância dos dados para cada um dos níveis, bem como as principais ferramentas estatísticas utilizadas nos níveis 4 e 5 do CMMI, e sua correspondência à filosofia Estatística do Seis Sigma.

	Nível 1	Nível 2	Nível 3	Nível 4	Nível 5
Medição	Coleta de dados e análise são <i>ad hoc</i> .	Dados são usados no planejamento e na gerência de projetos individuais.	Dados são coletados e usados em todos os processos definidos.	Definição e coleta de dados são padronizados através da organização.	DOE Dados são usados para avaliar e selecionar melhoria de processos.
			Dados são sistematicamente compartilhados entre projetos.	CEP Dados são usados para compreender o processo quantitativamente e estabilizá-lo.	6 σ

Figura 2: Correspondência entre CMMI e Seis Sigma: Resumo

De acordo com a Figura 2, o nível 4 apresenta o CEP (Controle Estatístico de Processo) como ferramenta estatística. Deming (1990) afirma que, para finalidades analíticas (tais como para melhoria de um processo da organização), as análises estatísticas não têm qualquer valor prático para se fazer a melhoria de um processo, a não ser que os dados sejam produzidos por um sistema (um processo) que esteja em um estado de controle estatístico. O primeiro passo no exame de dados estatísticos é, portanto, segundo Deming (1990), questionar o estado de controle estatístico que produziu esses dados. Tais conceitos são também aplicáveis à melhoria de processos de software, segundo reconhece Humphrey (1988), sendo que a maneira mais fácil de examinar os dados é colocá-los numa carta de controle, na ordem em que são produzidos, para se averiguar qual proveito pode ser tirado da distribuição dos dados ao longo da carta.

Todavia este artigo tem foco na Melhoria de Processo de Software. Isso significa que iremos tratar apenas do nível 5 do CMMI, e da ferramenta estatística DOE proposta para o alcance de tal nível.

Para que os delineamentos experimentais (DOE) possam ser realizados, afim de que uma empresa de software alcance o nível 5 do CMMI, conforme apresentado na figura anterior, os dados devem estar organizados de forma a possibilitar esta análise estatística dos mesmos.

A possibilidade de realização de DOE nos dados coletados é tratada neste artigo como a **organização estatística** dos dados. Organização esta provida pela COPE, a qual vincula o modelo PSM com conceitos Estatísticos.

A Correspondência PSM vs Estatística (COPE)

O Modelo PSM e a ferramenta PSM Insight (2013) tem sido um importante auxílio para as empresas de software que desejam realizar controle estatístico de processo (Nível 4 do CMMI) e melhoria contínua (Nível 5 do CMMI). Nesta ferramenta existe um conjunto de medições já identificadas e organizadas considerando boas práticas de desenvolvimento de software, adquiridas em experiências anteriores.

Todavia, para que uma empresa alcance estes níveis de qualidade não basta apenas uma boa organização dos dados no sentido de rastreabilidade dos mesmos, mas tem-se como fator mais importante a organização estatística dos dados, ou seja, um armazenamento inteligente dos mesmos, possibilitando a geração de futuros gráficos de controle (Nível 4 do CMMI – estabilização) e a construção de delineamentos experimentais (Nível 5 do CMMI - melhoria), dentre outras análises Estatísticas.

Os dados devem ser organizados pensando-se estatisticamente, e não apenas em facilidade de localização. Pois, na primeira forma de organização estes deixam de ser dados e tornam-se informações valiosas para o controle e melhoria de um processo produtivo.

A ideia desta correspondência então, é aproveitar o que o PSM Insight tem de melhor, ou seja, a organização da pré-identificação de medições baseadas em boas práticas, e mapear os conceitos teóricos desta ferramenta e modelo, para conceitos Estatísticos. Promovendo uma utilização em conjunto de ambas as abordagens: O modelo PSM e a Estatística.

As etapas desta correspondência podem ser detalhadas da seguinte forma:

Foram observadas no PSM Insight (2013) várias medições que podem ser consideradas em um processo de medição de um projeto de software. Conforme dito anteriormente, vinculado a essas medições existem características das mesmas. Observe Figura 3.



Figura 3: Características dos dados

Como mostrado na figura anterior, para um determinado dado, existem características (etiquetas) vinculadas ao mesmo. Características essas conhecidas no PSM como estruturas e atributos, conforme visto anteriormente.

Por meio de um estudo da representatividade estatística dessas características, as mesmas foram mapeadas estatisticamente para o conceito de fatores, conforme representado pela Figura 4.

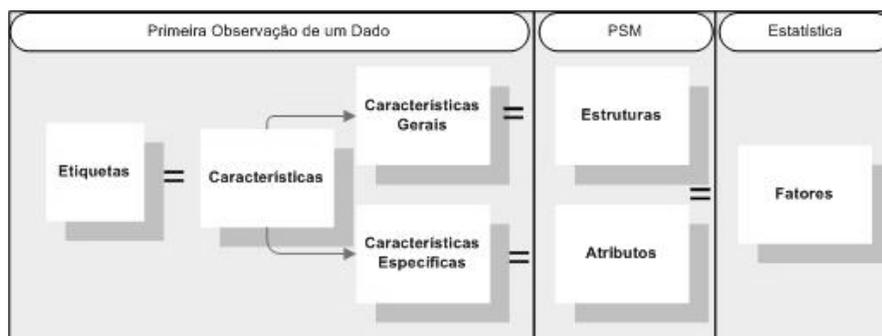


Figura 4: As Etiquetas de um dado e a visão do PSM e da Estatística

Uma observação importante a ser feita considerando essa figura é a grande quantidade de etiquetas vinculadas aos dados na Ferramenta PSM Insight (Figura 5).



Figura 5: Exemplo de uma medida no PSM e seus atributos e estruturas

Tendo em vista o conceito estatístico de fatores e níveis, definidos por Steel e Torrie (1980), temos a esquematização dos mesmos conforme Figura 6. Nesta figura percebemos que associados a um determinado fator, podemos ter vários níveis. Por exemplo, associado ao fator “Prioridade Requisito” podemos ter os níveis: Alta, Média e Baixa.

Fator 1	Fator 2	Fator 3
Nível A	Nível C	Nível E
Nível B	Nível D	Nível F

Figura 6: Fatores em um Delineamento Experimental

Neste mesmo contexto o conceito de tratamentos deve ser considerado, os quais são compostos seguindo a estrutura mostrada através da Figura 7. Ou seja, os tratamentos surgem a partir da combinação dos níveis dos fatores associados às medidas. Por exemplo, podemos imaginar o tratamento 1 como sendo formado pelos níveis alta, primeira e empresarial, dos respectivos fatores “Prioridade Requisito”, “Iteração” e “Tipo de Organização”.

Tratamentos			
1	Nível A	Nível C	Nível E
2	Nível A	Nível C	Nível F
3	Nível A	Nível D	Nível E
4	Nível A	Nível D	Nível F
5	Nível B	Nível C	Nível E
6	Nível B	Nível C	Nível F
7	Nível B	Nível D	Nível E
8	Nível B	Nível D	Nível F

Figura 7: Tratamentos em um Delineamento Experimental

Para a melhoria de um processo de desenvolvimento de software, deve ser realizado um estudo do mesmo por meio de um conceito estatístico denominado tratamento. Conforme visto no tópico sobre **definição de conceitos estatísticos**, os diferentes tratamentos ligados a uma determinada variável (o que é chamado de medida no PSM) são comparados, e por meio dessas comparações, conclusões para a melhoria do processo são encontradas.

Diante de todas as considerações anteriores, a Figura 8, a seguir, exemplifica um tratamento considerando a medida Número de Requisitos (da Figura 5), existente na Ferramenta PSM Insight (2013).

Tratamento			
1	Imprimir Boletim	Planejado	Empresa X
	Alta	Componente Bancário	Primeira
	Empresarial		Requisito de Sistema
Valor da Variável [Total de Requisitos] = 55			

Figura 8: Exemplo de Tratamento

Observa-se pela Figura 8 uma grande quantidade de fatores que caracterizam este tratamento, tornando-o complexo, no que se diz respeito à realização de um delineamento experimental. Esta dificuldade de interpretação é representada pela Figura 9.



Tratamentos								
1	2	3	4	5	6	7	8	9
10	11	12	13	14	15	16	17	18
19	20	21	22	23	24	25	26	...

Figura 9: Dificuldade de Interpretação

A figura representa a dificuldade de interpretação do dado '55', uma vez que o mesmo está ligado a um tratamento composto por um grande número de níveis (neste caso, 26).

Diante dessas primeiras observações da utilização da ferramenta PSM Insight (2013) - principalmente com relação à dificuldade de interpretação do significado dos dados - como ferramenta de medição e armazenagem de dados em um processo de software, seguindo a filosofia Estatística de um controle estatístico de processo, evidenciou-se a necessidade de uma correspondência entre conceitos do PSM para conceitos estatísticos (COPE).

Após etapas de estudo quanto à verificação da representatividade dos conceitos do PSM para termos Estatísticos chegou-se a conclusão esboçada através da Figura 10:

PSM	Estatística
"Issue" - Problema	Experimento
"Category" - Categoria	
"Measure" - Medida	
"Attribute" - Atributos	Fatores
"Structures" - Estruturas	
Combinação entre Atributos e Estruturas	Tratamentos
"Data Items" - Item de Dados	Variáveis Resposta

Figura 10: Correspondência entre PSM e Estatística

A partir da definição das medidas, seus atributos e estruturas relacionadas, temos a definição do que chamamos em Estatística de tratamentos de um experimento. O que em um delineamento experimental (DOE) em fatorial é dado por intermédio do cruzamento dos fatores.

Para que se tenha uma ideia do quanto seria difícil à interpretação dos dados considerando a organização baseada na rastreabilidade, podemos citar que em muitos casos, da forma como estão organizados os dados no PSM Insight, seria impossível realizar delineamentos experimentais. Isso porque as medições do PSM Insight estão ligadas a muitos atributos e estruturas, o que significa estatisticamente que em muitos experimentos existem fatores em excesso. Inviabilizando a realização de delineamentos experimentais e conseqüentemente a interpretação dos dados.

Para que um experimento em fatorial possibilite uma boa interpretação, o mesmo deve ter no máximo três fatores, o que não ocorre com a maioria das medições identificadas previamente na ferramenta PSM Insight.

A Figura 11, a seguir, mostra a seqüência de passos em que os termos de conceitos da ferramenta PSM Insight são utilizados estatisticamente.

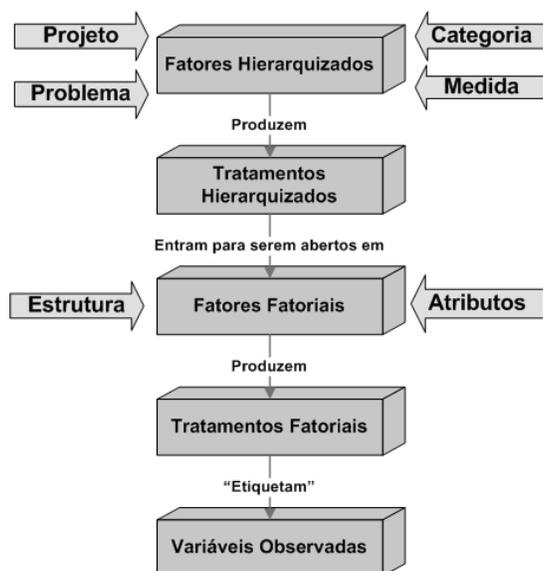


Figura 11: Metodologia de utilização dos termos do PSM *Insight* Estatisticamente

Interpretando a Figura 11 temos que:

- Fatores Hierarquizados: a partir da definição do projeto, problemas, categorias de medidas e medidas, pode-se dizer estatisticamente que os fatores hierarquizados são identificados.
- Tratamentos Hierarquizados: com as identificações realizadas anteriormente os tratamentos hierarquizados são delineados.
- Fatores Fatoriais: depois de definidos os tratamentos hierarquizados, o próximo passo é abrir cada um desses tratamentos, que na verdade correspondem em seu mais baixo nível às medidas do PSM Insight, em fatores fatoriais que os influenciam. As estruturas e os atributos são esses fatores.
- Tratamentos Fatoriais: por meio do cruzamento entre os fatores fatoriais são identificados os tratamentos fatoriais.
- Variáveis Observadas: os itens de dados são as variáveis estatisticamente falando. O que realmente é medido no PSM Insight (2013), e produz números para serem analisados. Essas variáveis são identificadas, ou etiquetadas, por meio dos tratamentos fatoriais.

Conclusão

Por meio da correspondência proposta conclui-se que com a organização Estatística dos dados na ferramenta PSM Insight (2013), as informações obtidas dos mesmos se tornam facilitadas. Dados foram simulados e delineamentos experimentais realizados na finalidade de confirmar esta facilitação.

Os delineamentos experimentais, base para identificação de melhorias ao processo de desenvolvimento de software, e ponto principal a ser considerado no nível 5 do CMMI pelas empresas de software, podem ser realizados com maior facilidade.

Organizar os dados estatisticamente é mais do que etiquetar dados. As informações que verdadeiramente importam com relação aos dados obtidos não se referem a *onde* eles estão (rastreabilidade), mas sim em *o que* eles representam estatisticamente.

Referências bibliográficas

DEMING, W. E. *Qualidade: a revolução da administração*. Editora: Marques – Saraiva, Rio de Janeiro, 1990.

HUMPHREY, W. S. *Characterizing the software process*. IEEE Software, 1988.

MCGARRY, J.; CARD, D.; JONES, C.; LAYMAN, B.; CLARK, E.; DEAN, J.; and HALL, F. *Practical Software Measurement – Objective Information for Decision Makers*. Addison Wesley, 2002.

OLIVEIRA, M. S. *Qualidade de Processo Software: Medição e Análise*. Editora UFLA, 2006.

PARK, R.E.; GOETHERT, W.B.; FLORAC, W.A. *Goal-Driven Software Measurement*. In: A Guidebook, CMU/SEI-96-HB-002, Pittsburgh, PA, Software Engineering Institute, Carnegie Mellon University, 1996.

PAULK, M. C.; WEBER, C. V.; CURTIS, B.; CHRISISS, M. B. et al. *The capability maturity model: guidelines for improving the software process*. Addison Wesley, 1995.

PSM. *PSM Insight Software, Version 4.2.3*. Disponível em <http://www.psmc.com>. Acesso em 03/03/2013.

SANTOS, R. V. *Definição de um processo de medição de software baseado em Seis Sigma e CMMI*. Lavras, UFLA, 2007. 177 p. (Dissertação de Mestrado)

SNEE, R. D. *Impact of Six Sigma on quality engineering*, Quality Engineering, 2000.

STEEL, R.G.D.; TORRIE, J.H. *Principles and procedures of statistics: a biometrical approach*, 2.edition, New York : McGraw-Hill, 631p, 1980.

USEVICIUS, L. A. *Implantação da metodologia Seis Sigma e aplicação da técnica estatística projeto de experimentos na resolução de problemas e otimização de processos de fabricação*. (Dissertação de Mestrado Profissional em Engenharia de Produção) - Universidade Federal do Rio