

UMA METODOLOGIA DE CHI-QUADRADO APLICADA AO CONTROLE DE DESVIOS NO PLANEJAMENTO DE PROJETOS

Denis Ávila Montini¹, Danilo Battaglia², Gustavo Ravanhani Matuck³, Paulo Marcelo Tasinaffo⁴, Luiz Alberto Vieira Dias⁵, Alessandra Ávila Montini⁶, Adilson Marques da Cunha⁷

Abstract — This paper describes a hybrid-Chi-squared methodology for Project Plan deviations identification, including those for multiple deviations in a Project Plan system. The sampling methodology was used to analyze similarity, together with a statistical optimization methodology for Chi-squared, to select the methodology used to forecast data in the Project Plan phase. The created sampling methodology was utilized for Project Plan activities learning, in order to detect deviations. The hybrid-Chi-squared methodology was employed by empirical results. The results have shown that the selection strategy identify the best forecasting approach between two different proposals, aiming to provide the Project Plan forecasting process, presenting the best precisions based on the diagnose deviations detection, within this hybrid-Chi-squared methodology.

Keywords – Chi-squared statistical methodology, Project Estimation forecasting, sampling methodology, and Estimates driven models.

Resumo — Este artigo descreve uma metodologia híbrida Chi-Quadrado para a identificação de desvios em Planejamento de Projeto, incluindo aqueles para os desvios múltiplos em um sistema de Plano de Projeto. Uma metodologia de amostragem multivariada foi usada para analisar semelhança, juntamente com uma metodologia de otimização estatística do Chi-Quadrado para selecionar a metodologia utilizada dados de previsão em fase de Plano de Projeto. A metodologia de amostragem foi utilizada para atividades do projeto Plano de Aprendizagem, a fim de detectar desvios. A metodologia híbrida Chi-Quadrado foi empregada por meio de testes empíricos. Os resultados mostraram que uma estratégia de seleção para identificar a melhor abordagem de previsão entre duas propostas diferentes. A aplicação da metodologia híbrida de Chi-Quadrado teve por objetivo fornecer ao Projeto um processo de previsão, a fim de realizar um plano de projeto mais realista e apresentar previsões mais precisas e baseadas na detecção de desvios diagnóstico.

Palavras-chave – Chi-Quadrado, Detecção de desvio de Projeto e Estimativas orientadas por modelos.

INTRODUÇÃO

Com a finalidade de melhorar algumas técnicas de previsões convencionais e assertividades de Projeto, como por exemplo, análise de Pontos por Função (APF) [1], uma metodologia foi desenvolvida com o objetivo de detectar os desvios de projeto, para a aplicação sistematizada do conceito do Chi-Quadrado.

Esta metodologia foi aplicada, no contexto de linhas de produção de *software* [2], utilizando uma forma de planejamento estabilizada e operacional. Devido as limitações das linhas de produção, a metodologia de Chi-Quadrado [3] monitorou uma sequência de projetos semelhantes, na mesma linha e com a mesma equipe, para evidenciar desvios entre planejamentos e execuções, em um mesmo Intervalo de Confiança (IC).

Nesse contexto, um algoritmo que implementa o teste estatístico de Chi-Quadrado foi computacionalmente desenvolvido em MATLAB®, para automatizar o processo de identificação de desvios de intervalos de tempo, entre o planejado e o executado em projetos [4][5].

A proposta de amostragem que incorporou o teste estatístico de Chi-Quadrado foi utilizada [4], pois os parâmetros desta metodologia proporcionaram o uso do teste estatístico Chi-Quadrado [3][5][6]. Esta metodologia foi empregada em uma atividade de monitoramento, durante o andamento de projetos [2][7].

Neste trabalho, uma estratégia de seleção de padrões de projeto foi proposta e implementada, por meio da utilização de um algoritmo de amostragem [4], em MATLAB®, identificando melhores padrões de projeto e utilizando o Chi-Quadrado nos desvios de planejamento [3][5][6].

APLICANDO O CHI-QUADRADO À SISTEMAS

O gerenciamento de desvios em projetos pode ser realizado de diversas formas e utilizando diferentes abordagens como, por exemplo, a estratégia de organizar dados a fim de viabilizar sistematicamente o teste estatístico

¹ Denis Ávila Montini, Msc. Doutorando ITA, São José dos Campos, SP, Brasil, denisavilamontini@yahoo.com.br

² Danilo Battaglia, Aluno de Pós Graduando na FIAP, São Paulo, SP, Brasil, danilo.battaglia@terra.com.br

³ Gustavo Matuck, Msc. Doutorando ITA, São José dos Campos, SP, Brasil, gmatuck@ita.br

⁴ Paulo Marcelo Tasinaffo, Professor Doutor do ITA, São José dos Campos, SP, Brasil, tasinaffo@ita.br

⁵ Luiz Alberto Vieira Dias, Professor Doutor do ITA, São José dos Campos, SP, Brasil, vdias@ita.br

⁶ Alessandra de Ávila Montini, Professora Doutora da FEA-USP, São Paulo, SP, Brasil, amontini@usp.br

⁷ Adilson Marques da Cunha, Professor Doutor do ITA, São José dos Campos, SP, Brasil, vdias@ita.br

do Chi-Quadrado. A sistematização deste algoritmo não é uma tarefa trivial, quando aplicada às Linhas de Produção [2][7].

A metodologia desenvolvida e adaptada com base na Chi-Quadrado combinou informações qualitativas e quantitativas, utilizando cálculos da estatística descritiva. Esta abordagem foi aplicada especificamente no controle e monitoramento da variabilidade dos desvios encontrados em projetos de *software* no ambiente de produção [2][7].

Para viabilizar a implementação do Chi-Quadrado, utilizou-se, como hipótese nula (H_0), as diferenças coincidentes observadas. Portanto, a hipótese H_0 assumiu que o analista quis rejeitar as diferenças coincidentes observadas, considerando o maior grau de significância possível [5][6].

Por outro lado, a hipótese alternativa (H_1), significou o oposto da hipótese nula. Portanto, o aceite das diferenças coincidentes observadas [5][6]. Ambos testes estatísticos, H_0 e H_1 , viabilizaram a determinação da aceitação ou da rejeição da hipótese nula, H_0 .

Neste caso, o teste de hipótese foi conduzido em seis etapas: (i) Determinação do Nível de Significância do teste; (ii) Escolhas estatísticas de uma distribuição conhecida, a partir de H_0 ; (iii) Construção de uma Região Crítica e de um Nível de Significância; (iv) Obtenção de um valor estatístico; e (v) Rejeição da hipótese nula, H_0 .

Nesta pesquisa, a técnica estatística de Chi-Quadrado de Pearson [3][5][6] foi realizada, por meio de comparações entre: as Frequências Observadas (*Observed Frequency* - OF), a partir de um grupo de testes de *software*, a fim de identificar as variáveis categóricas; e as Frequências Esperadas (*Expected Frequency* - EF), obtendo-se assim os indicadores primários.

O Algoritmo desenvolvido considerou que os dados

variáveis foram pareados, ou seja, possuíam relação direta entre eles. As variáveis escolhidas para o teste seguiam a distribuição Normal. Neste caso, a EF foi constatada, a *posteriori*, e confrontada no experimento.

No cálculo estatístico do processo foi utilizada a fórmula descrita na Figura 1.

$$\chi^2_{calc} = \sum_{i=1}^r \sum_{j=1}^s \frac{(OF_{ij} - EF_{ij})^2}{EF_{ij}}$$

Figura 1 - Fórmula do Chi-Quadrado.

A técnica para a estatística de Chi-Quadrado de Pearson, como em [3][5][6], baseou-se na magnitude da diferença entre OF e EF. Portanto, para esta diferença do Chi-Quadrado foi atribuída em cada fase do ciclo de desenvolvimento de *software*. Inicialmente, a técnica estatística de Chi-Quadrado de Pearson foi calculada por encontrar a diferença entre OF e EF para cada resultado possível, como em [3][5][6].

A METODOLOGIA PROPOSTA PARA A IDENTIFICAÇÃO DE DESVIOS

A metodologia de identificação de desvios em projetos foi constituída por meio de um novo modelo híbrido Chi-Quadrado pois, além de combinar as variáveis pareadas, esta se utilizou de alguns indicadores estatísticos descritivos [3][4].

A metodologia de cálculo do Chi-Quadrado proposto foi composta por sete passos resumidos na Figura 2:

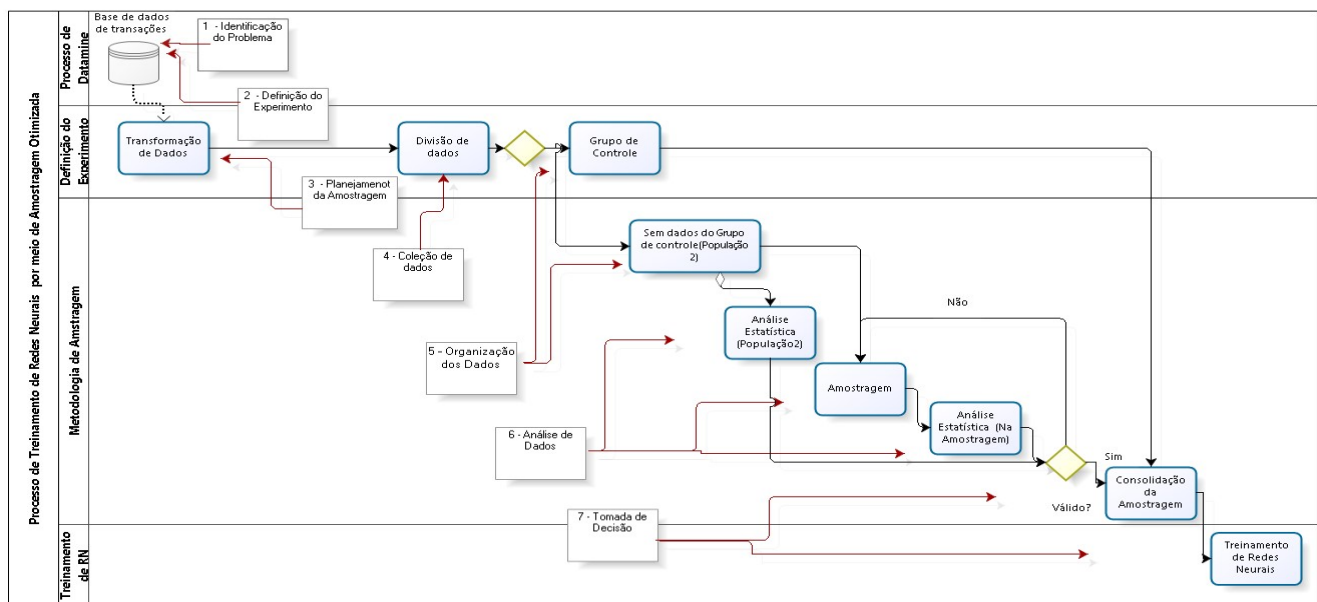


Figura 2 – O diagrama da metodologia Chi-Quadrado aplicado a detecção de desvios em Projetos. Adaptado de Montini [4].

Passo 1 - A identificação do problema é realizada por meio da restrição de âmbito da experiência;

Passo 2 - A definição do experimento consiste em uma pesquisa para definir a forma de dados e os indicadores descritivos que devem ser registradas por meio de estatística, a fim de obter os resultados tabulados. A pesquisa também requer um alfa (α), Grau de Especificação [3], com a sua identificação de precisão desejada;

Passo 3 - O Planejamento do Chi-Quadrado define uma configuração experimental composta pela: (i) criação de uma unidade experimental, contendo medições de um resumo estatístico; (ii) seleção das variáveis em análise e suas formas de medição; e (iii) criação de indicadores expressos por meio de medições de resumo, de modo a garantir os tratamentos comparativos;

Passo 4 - A coleta de dados consiste de um estudo para extração e validação de experiências. Neste passo, considera-se aceita a coleta, se os parâmetros obtidos da amostra estiverem dentro do Intervalo de Confiança (IC) dos parâmetros da população. Neste quarto passo, também é criado um procedimento recursivo para a seleção de novos testes do Chi-Quadrado;

Passo 5 - A etapa de Organização de Dados converte os

dados para um formato diferente, de acordo com uma definição de Planejamento do Chi-Quadrado;

Passo 6 - A etapa de Análise de Dados verifica se os objetivos foram alcançados, usando a abordagem para o Chi-Quadrado; e

Passo 7 - A tomada de decisão considera os resultados da análise de dados para proporcionar informação de decisão, de acordo com a identificação do problema proposto. A Figura 2 mostra um diagrama com o leiaute proposto para uma forma de se realizar diagnósticos por meio do uso de uma metodologia de desenvolvimento de *software* [4].

APLICANDO A METODOLOGIA DO CHI-QUADRADO EM CONTROLE DE PRODUÇÃO

A metodologia do Chi-Quadrado aplicada ao Planejamento e Controle de Projetos foi sistematizada e constituída de: (i) um planejamento de projetos [7]; (ii) a utilização de controles estatísticos [3][5][6]; (iii) uma análise multivariada; e (iv) uma utilização de testes estatístico, utilizando-se dos cálculos de Chi-Quadrado. A seguir, a Figura 3 ilustra a metodologia proposta.

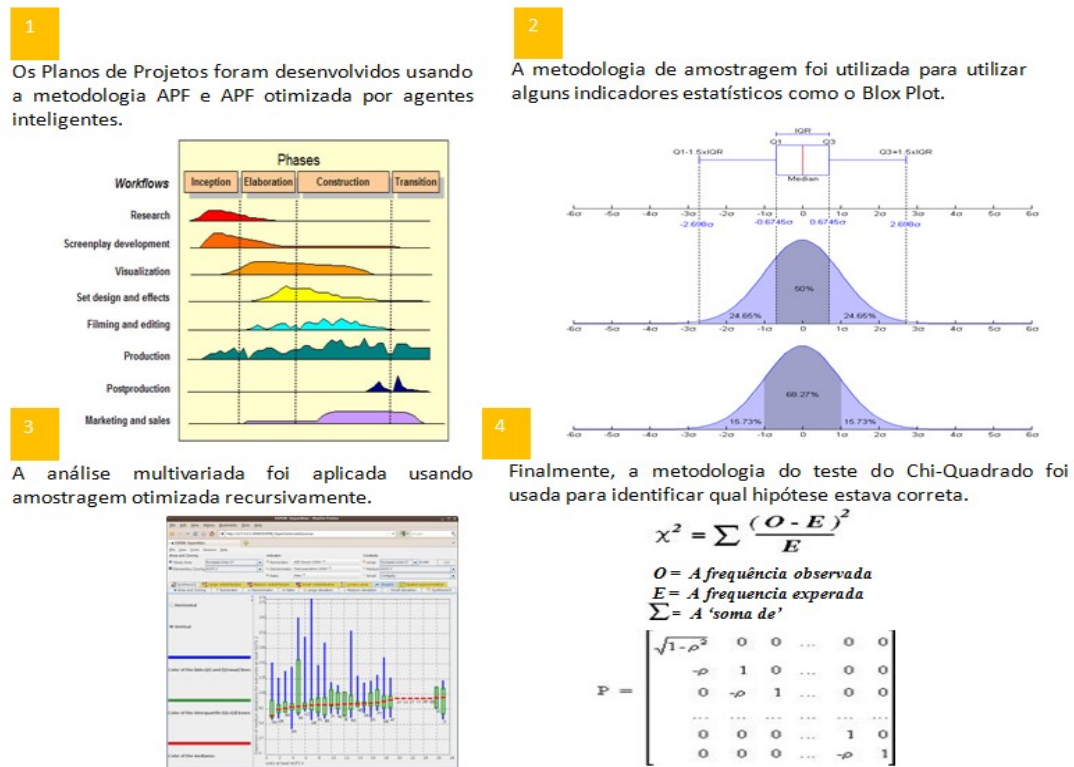


Figura 3 - A metodologia do Chi-Quadrado para utilização em Planejamento e Controle de Projetos.

DADOS ESTRATIFICADO OTIMIZADOS PELO PROCESSO DE SELEÇÃO

A metodologia proposta foi aplicada em 18 projetos diferentes. Os indicadores obtidos a partir de cada um desses projetos, utilizando a estatística descritiva foram: (i) Média; (ii) Média Ponderada; (iii) Mediana; (iv) Variância; (v) Desvio Padrão; (vi) Amplitude; e (vii) Coeficiente de Variação.

Esses parâmetros foram utilizados para a população. Depois disto, o algoritmo previu, recursivamente, pesquisas, a fim de encontrar, na base de dados de Projetos, um melhor parâmetro de Chi-Quadrado.

Neste caso, o melhor parâmetro de Chi-Quadrado foi obtido, por meio de intervalos indicadores estatísticos e a partir da base de dados de Projetos anteriormente descrita.

Desses cruzamentos de dados entre a população e os intervalos do indicador de Chi-Quadrado, obtido da base de dados, foi selecionado o melhor Indicador Chi-Quadrado que continha o menor valor.

Tabela 1 – Média histórica de Projetos.

Project Plan Intervals	Mínimo	Parâmetro	Máximo
Média Aritmética	0,2106	0,2343	0,2577
Média Ponderada	0,2106	0,2343	0,2577
Mediana	0,1250	0,1391	0,1530
Variância	0,0391	0,0438	0,0481
Desvio Padrão	0,1877	0,2092	0,2301
Amplitude	0,4251	0,4738	0,5212
Coeficiente de Variação	80,2096	89,2897	98,2187

A metodologia proposta e implementada para Controle de Projetos, utilizando Chi-Quadrado nos desvios entre OF e EF, foi aplicada em 18 projetos, utilizando os dados do projeto que obteve o menor desvio, segundo o cálculo do Chi-Quadrado.

A Tabela 2 contém os dados do Projeto que obteve o menor desvio de dados em relação a média histórica de Projetos, bem como os resultados obtidos com menor Chi-Quadrado.”

Após a utilização dos dados de Planejamento, o Chi-Quadrado foi novamente utilizado para verificar se houve uma melhora e uma diminuição nos desvio entre o OF e EF.

Tabela 2 – Dados do Novo Projeto.

Project Plan Intervals	Mínimo	Parametro	Máximo
Média Aritmética	0,2141	0,2834	0,3542
Média Ponderada	0,2141	0,2834	0,3542
Mediana	0,1243	0,1671	0,2088
Variância	0,0613	0,0780	0,0975
Desvio Padrão	0,2147	0,2792	0,3490
Amplitude	0,4743	0,6180	0,7725
Coeficiente de Variação	75,2483	98,5313	123,1641

ANALISANDO OS RESULTADOS DE H0

O método de controle de amostragem constituiu-se por meio de uma camada de entrada, contendo 18 projetos, e seus dados estruturados em um sistema de banco de dados.

Nos três conjuntos de dados históricos analisados, a distribuição de desvios e de atividades não continha os registros de desvios. Neste caso, todos os dados usados nas duas hipóteses, H_0 e H_1 , foram divididos em dados da população contendo informações de 17 projetos e mais o planejamento de um projeto adicional, inteirando os 18.

A hipótese H_0 foi confirmada, para avaliar a qualidade do planejamento, utilizando a técnica estatística de Chi-Quadrado de Pearson [3][5][6]. Neste cenário, foi possível afirmar que a hipótese H_0 foi obtida a partir de dados experimentais, utilizando a fórmula da Figura 1 e considerando o erro de tipo $\alpha = 0,05$.

A constatação de que a H_0 foi embasada nos dados obtidos dos indicadores proporcionou a obtenção do Planejamento e Controle para o décimo oitavo projeto, obtendo-se um Intervalo de Confiança (IC) igual a 0,05.

CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho, uma metodologia proposta de Chi-Quadrado para a seleção empírica de amostragem em um sistema de banco de dados de Projetos foi concebida, implementada e avaliada, a partir de dois grupos de experimentos realizados.

Nos testes de hipótese realizados, o nível de significância do teste Chi-Quadrado adotado foi utilizado para determinar que a probabilidade de aceitação de H_0 com níveis de significância médio foi igual a 10% [5][7].

A abordagem da metodologia proposta neste artigo foi aplicada em atividades de Planejamento e Controle de Projetos com estruturas e maturidades organizacionais previamente definidas, para vender diversos projetos de mesmas características.

A utilização da metodologia descrita neste trabalho foi obtida, a partir de uma área de medição e análise capaz de manter os dados necessários para os processos de inferências estatísticas.

A principal vantagem constatada a partir da aplicação da metodologia proposta utilizando o Chi-Quadrado foi a reutilização de dados de Planejamento e Controle de Projetos de linhas anteriores.

O tempo de planejamento e a inicialização destes projetos foram reduzidos em 90%. Neste caso, foram atingidos e identificados desvios semelhantes de Intervalo de Confiança (IC), iguais a 5%, no décimo oitavo projeto. Este projeto se encontrava dentro do Intervalo de Confiança (IC) médio proposto igual a 10%.

No entanto, vale a pena ressaltar que, a maior dificuldade encontrada durante esta pesquisa foi o acesso a dados confiáveis de uma linha de produção específica.

Neste caso, constatou-se que os desvio nos IC dos Chi-

Quadrados encontrados ocorreram, principalmente, devido a falta de conhecimento científico a respeito: (i) do produto; (ii) do ciclo de vida de desenvolvimento; (iii) dos processos; (iv) dos padrões de qualidade; (v) das metodologias envolvidas; (vi) das tecnologias envolvidas; e (vii) dos desempenhos dos participantes dos projetos.

Em relação as metodologias Chi-Quadrado aplicadas ao Controle de Projetos, foi possível prover elementos de controle, como IC mais precisos, que viabilizassem Planejamentos com previsões mais precisas para diminuir problemas durante a execução de projetos.

CONSIDERAÇÕES E RECONHECIMENTO

A manutenção da infraestrutura de pesquisa em engenharia de *software* tem sido um ponto importante para a continuidade de projetos de pesquisa e desenvolvimento em todo o mundo. Instituições e grupos de pesquisa, como o Instituto Tecnológico de Aeronáutica (ITA) e o Grupo de Pesquisa de Engenharia de *Software* (GPES) têm contribuído para o avanço desta obra especializada.

Os autores agradecem a empresa 2RPNet, uma empresa Brasileira, e ao ITA, por suas contribuições, cooperações e apoios, durante todas as fases iniciais desta pesquisa.

REFERÊNCIAS

- [1] ALBRECHT, ALLAN J. Function Points as a Measure of Productivity, Act as do 53 rd meeting of GUIDE International Corp., Guidance for users of integrated data processing equipment conference, Dallas, 1981.
- [2] MONTINI, D. Á.; TASINAFPO, P. M.; MONTINI, A. A.; DIAS, L. A. V.; CUNHA, A. M.; “Um Meta-Algoritmo para Otimização de Planejamento em Linha de Produção de software”. In: VIII International Conference on Engineering and Computer Education - ICECE 2013.: Luanda, Angola, 3-6/Março/2013.
- [3] PLACKETT, R. L.; “Karl Pearson and the Chi-Squared Test,”in: International Statistical Review, International Statistical Institute (ISI);1983 vol. 51; pp. 59-72. Disponível em <http://www.jstor.org/discover/10.2307>, Junho de 2013.
- [4] MONTINI D. Á; MATUCK G.R , Ribeiro A. L.P D. A.; DIAS L. A. V; CUNHA, MONTINI A. M. da., , A. A., “A Sampling Diagnostics Model for Neural System Training Optimization” 101h International Conference on Information Technology New Generations (ITNG’13), Las Vegas, IEEE Computer Society, Abril de 2013.
- [5] STEVENSON, W., J; ISSELHARDT B. J.; “Study Guide to Accompany Business Statistics: Concepts and Applications”, Segunda Edição Harper & Row. ISBN 0060464410, 1985.
- [6] MENDENHALL, W.; “Introduction to Probability and Statistics”, Sexta Edição Ed. Editora PWS, pp. 371-381; 493-494; 498-503; A8. 442137, 1983.
- [7] MONTINI, D. A, “Modelo de indicadores de risco para o orçamento de componentes de software para célula de manufatura”. Dissertação de Mestrado em Engenharia de Produção, Universidade Paulista, São Paulo, Brasil, 2005.