

Uma Abordagem Simplificada para Experimentação de Artefatos em Trabalhos de Conclusão de Curso*

Felipe Santos, Kele Belloze, Fabio Jr,
Eduardo Bezerra, João Quadros, Eduardo Ogasawara

Escola de Informática & Computação – CEFET/RJ
Av. Maracanã, 229, Rio de Janeiro, RJ, 20271-110

{fabio.paschoal, ebezerra, jquadros, eogasawara}@cefet-rj.br,
kelebelloze@gmail.com

Abstract. *Final year projects (FYP) are a common requirement for students in Computer Science and Information Technology courses. These projects commonly develop artifacts that tackle some problem related to software development. Nevertheless, it is also very common to develop artifacts that are not evaluated. One of the main reasons for that is the lack of practice of students in doing experimentation. This paper aims at solving this issue by establishing a simplified approach for the production of experiments with artifacts resulting from FYP. Our approach involves using evaluation questionnaires and computational tools to convey the statistical analysis of the results. We present a case study on the application of this approach in two FYPs. This case study showed that the involved students could rapidly accomplish the statistical evaluation of the artifacts produced in their FYPs.*

Resumo. *Trabalhos de Conclusão de Curso (TCCs) são um requisito comum para estudantes em cursos de Ciências da Computação e de Tecnologia da Informação. Nesses trabalhos, geralmente são desenvolvidos artefatos de software como produtos do problema proposto. Entretanto, em muitos TCCs não são produzidas avaliações experimentais dos artefatos produzidos. Um dos principais motivos que levam a isso é a falta de prática dos alunos em experimentação. Este trabalho visa a apoiar essa demanda ao estabelecer uma abordagem simplificada para a realização de experimentos em artefatos de software gerados em TCCs. Nossa abordagem compreende o uso de questionários de avaliação e de ferramentas computacionais para facilitar a análise estatística dos resultados. Apresentamos um estudo de caso da aplicação dessa abordagem em dois TCCs. Esse estudo mostrou que os alunos envolvidos puderam realizar rapidamente a avaliação estatística dos artefatos produzidos em seus TCCs.*

1 Introdução

Trabalhos de conclusão de cursos (TCC) em cursos de Ciência da Computação ou de Tecnologia da Informação são comumente produzidos por um ou dois alunos e sua realização normalmente tem duração de um ou dois semestres letivos. Neles, há a figura do professor orientador, cuja função é ajudar o aluno na definição do escopo e no processo de desenvolvimento do projeto. Os TCCs de computação, na maior parte dos

*Os autores agradecem ao CNPq e à FAPERJ pelo financiamento parcial do trabalho.

casos, envolvem a elaboração de um ou mais artefatos de software. Essa elaboração abrange diversas tarefas: pesquisa bibliográfica, definição do problema, escopo, modelagem e desenvolvimento da solução. Os produtos de um TCC compreendem a elaboração do(s) artefato(s) propriamente dito(s), de uma documentação e finalmente uma apresentação do projeto perante uma banca avaliadora.

Não raramente, percebe-se que alguns alunos se deparam tanto com dificuldades técnicas, pela falta de experiência prática, quanto com dificuldades de gerência para realizar o trabalho em grupo. Apesar destes desafios, muitas ideias e artefatos interessantes são elaborados nos TCCs. Em diversas vezes, entretanto, não são apresentadas, nesses TCCs, avaliações das soluções propostas. Um dos maiores empecilhos para isto é a falta de experiência dos alunos em realizar experimentos [Boyle et al. 2014].

Este trabalho descreve uma abordagem simplificada para produção de experimentos para avaliação de artefatos de software gerados em TCCs [Juristo and Moreno 2010, Wohlin and Regnell 2012]. Essa abordagem é composta por questionários de avaliação e por ferramentas para realização de análise estatística dos resultados. O objetivo da abordagem é fazer com que os alunos ganhem experiência em experimentação. A motivação para a definição dessa abordagem surgiu no âmbito dos cursos de graduação e do ensino médio-técnico da Escola de Informática & Computação do CEFET/RJ ao observar as dificuldades e a falta de experiência dos alunos da instituição em avaliar os artefatos produzidos em seus TCCs.

Não faz parte do escopo do presente trabalho cobrir todo espectro de avaliações possíveis. Apesar disto, a proposta é importante, principalmente em termos educacionais na área do ensino de computação. Isso porque frequentemente o profissional recém-formado se depara com o problema de avaliar componentes ou artefatos de software produzidos e neste sentido falta um direcionamento ou ponto de partida ou até mesmo a prática, em termos de ferramental, sobre como postular hipóteses e proceder experimentos de avaliação.

Além desta introdução, este trabalho é constituído por mais quatro seções. Na seção 2 é apresentado um resumo das principais formas de experimentação de artefatos em computação. A seção 3 apresenta a abordagem simplificada. A seção 4 descreve um estudo de caso de aplicação da abordagem proposta. A seção 5 apresenta as conclusões.

2 Experimentação de Artefatos

Seja qual for o artefato, avaliá-lo traz informações ou percepções adicionais a respeito dele. O resultado prático das avaliações pode estar relacionado a uma redução de custos de desenvolvimento, ao direcionamento de investimentos em novas funcionalidades ou até mesmo à produção de novos artefatos. Um dos mais importantes aspectos para o sucesso destas avaliações está no seu planejamento [Juristo and Moreno 2010]. Esta seção apresenta uma visão geral da temática relacionada à avaliação de artefatos. A seção 2.1 apresenta uma nomenclatura geral em experimentação. A seção 2.2 apresenta a forma de coleta de dados. A seção 2.3 descreve as análises estatísticas comumente realizadas nas avaliações. Finalmente, a seção 2.4 apresenta trabalhos relacionados.

2.1 Nomenclatura para Avaliação

Uma avaliação pode ser executada de várias maneiras. Porém, para a melhor compreensão dos diferentes tipos de avaliações, é importante clarificar a terminologia

comumente utilizada na área de experimentos de avaliação de artefatos [Juristo e Moreno 2010]. Em termos estruturais, têm-se os seguintes conceitos: *unidade experimental*, *sujeitos* e *replicação*, enquanto que, em termos de entradas e saídas, têm-se as *variáveis de resposta*, os *parâmetros* e *fatores*.

No que tange aos aspectos estruturais, compreende-se por *unidade experimental* o artefato que está passando por avaliação. Na área de computação, normalmente são software ou componentes. Os *sujeitos* são os responsáveis por executar a avaliação. Os sujeitos podem ser os usuários finais ou os desenvolvedores, dependendo do que está sendo avaliado. Cabe salientar que os sujeitos costumam ter grande influência na avaliação já que suas experiências e a maneira como eles lidam com os problemas costumam variar muito [Juristo and Moreno 2010]. Em função disto, costuma-se qualificar os sujeitos para que as análises possam refletir estas diferentes experiências. Finalmente, compreende-se por *replicação*, o número de vezes que a avaliação foi executada em diferentes unidades experimentais ou por diferentes sujeitos. A recorrente obtenção do mesmo resultado a partir de múltiplas replicações fundamenta a boa qualidade do experimento realizado.

No que se refere às entradas e saídas, entende-se por *variável de resposta* o resultado da avaliação. Os *parâmetros* compreendem as características do artefato as quais não se pretende mudar ao longo da avaliação. Isso é feito para se evitar a geração de oscilações nas variáveis de resposta. Finalmente, os *fatores*, em oposição aos parâmetros, são as características do artefato que podem influenciar na variável de resposta. Comumente as avaliações têm como objetivo medir o efeito dos fatores na variável de resposta. Em linhas gerais, fatores são as características da avaliação que são variadas propositalmente com o intuito de medir o impacto dessa variação sobre a variável de resposta.

2.2 Coleta de Dados

A coleta de dados pode ser obtida de modo direto ou indireto. Quando a coleta é direta, o próprio artefato produz os dados para análise. Quando a coleta é indireta, esta pode ocorrer por meio de medições sobre o uso do artefato, como, por exemplo, consumo de memória e duração de execução, ou coletadas a partir da percepção do sujeito a respeito do artefato. Neste último caso, é muito comum que se faça uso de questionários de avaliação [Sjoberg et al. 2005].

Os questionários podem incluir perguntas relacionadas à usabilidade dos artefatos, aspectos qualitativos dos sujeitos e aspectos relacionados aos artefatos. A parte referente a usabilidade corresponde a um conjunto de perguntas consolidadas que variam de acordo com a técnica escolhida. Uma das técnicas mais difundidas e consolidadas é o *System Usability Scale* (SUS) [Tullis and Albert 2008], que mede a concordância do sujeito, em uma escala de um a cinco, acerca de dez afirmações que produzem uma pontuação de usabilidade. Já a parte referente aos aspectos relacionados aos artefatos normalmente são formulados de modo que se formulem hipóteses para serem avaliadas estatisticamente.

2.3 Análises Estatística dos Resultados

No contexto de artefatos, a análise estatística objetiva produzir uma conclusão objetiva sobre algum artefato. A fundamentação teórica da análise estatística está na Inferência Estatística [Walpole 2012]. Esta é uma área da Estatística bem consolidada, bastante ampla e que apresenta diversas técnicas. O uso de uma dessas técnicas depende da

variável de resposta escolhida, da forma pela qual as amostras são obtidas e da maneira pela qual são conduzidas as avaliações.

Em linhas gerais, a análise estatística consiste em avaliar os fatores e relacioná-los às variáveis de resposta. Neste contexto, são comuns dois tipos de análise. A primeira, denominada análise em uma amostra, consiste em conferir se um determinado conjunto de fatores produz uma variável de resposta próxima ao desejado, *i.e.*, se um determinado fator X produz um artefato A com qualidade q . A segunda, denominada de análise em duas amostras, consiste em comparar as variáveis de resposta produzidas a partir de diferentes conjuntos de fatores. Aqui o intuito é averiguar se um fator X , que produz um artefato A com qualidade q_X é melhor, igual ou pior do que um fator Y ao produzir o mesmo artefato A com qualidade q_Y .

Avaliações que envolvem análise estatística podem ser quantitativas ou qualitativas ordinárias. As avaliações quantitativas medem grandezas diretamente mensuráveis, tendo um aspecto bem objetivo. Como exemplo, considere os tempos de execução de dois ou mais algoritmos. Outro exemplo são as quantidades de defeitos encontrados em duas ou mais versões de um artefato. Por outro lado, as avaliações qualitativas ordinárias estão relacionadas com a percepção do sujeito sobre a unidade experimental acerca da variável de resposta avaliada. Por exemplo, é possível que usuários em potencial avaliem qualitativamente a usabilidade da interface e a facilidade de uso dos artefatos construídos. Avaliações qualitativas inerentemente tendem a ter um aspecto subjetivo. Análises quantitativas costumam ser paramétricas, enquanto que as qualitativas ordinárias costumam ser não paramétricas [Walpole 2012]. Além disto, é comum a realização de avaliações que envolvem tanto dados quantitativos como também qualitativos.

Em termos práticos, considerando-se os cenários apresentados, o trabalho de avaliação consiste em postular uma hipótese nula (H_0) e uma hipótese alternativa (H_1). A partir de uma probabilidade p , definida *a priori* (adota-se, comumente, o valor de 5%), procura-se verificar se é possível refutar a hipótese nula. Quando isto ocorre, a hipótese alternativa é considerada verdadeira. Pode-se retomar o exemplo das técnicas X e Y previamente mencionadas. Neste exemplo, pode-se estabelecer que a hipótese H_0 é $q_X = q_Y$. Para formular a hipótese alternativa H_1 , pode-se adotar testes bilaterais ou unilaterais. No teste bilateral, H_1 é $q_X \neq q_Y$. No teste unilateral, H_1 pode ser tanto $q_X > q_Y$ quanto $q_X < q_Y$, dependendo do interesse do aluno.

2.4 Trabalhos Relacionados

O processo de aprendizado de experimentação e de domínio de estatística apresenta-se como um desafio para diversos alunos. Em termos de formulação dos experimentos, existe um conjunto de livros e sínteses na área de experimentação voltados à computação, principalmente quando relacionados à engenharia de software [Juristo and Moreno 2010, Sjoeborg et al. 2005]. Na área de Estatística, os livros textos compreendem um bom referencial [Walpole 2012]. Entretanto, a realização adequada de experimentos e a correspondente análise estatística dos resultados demandam diversas habilidades, tais como capacidade de abstração, habilidade de raciocínio lógico, análise crítica, formulação de hipóteses e interpretação de resultados. Se este processo não for bem direcionado, torna-se uma atividade desagradável e que gera muito ansiedade para os alunos. Boyle et al. (2014) apresenta uma síntese de diversas iniciativas e de práticas de estatística para tornar o aprendizado de experimentação mais atraente por meio de jogos educacionais, simuladores e outras abordagens mais lúdicas.

3 Abordagem Simplificada para Experimentação de Artefatos

A abordagem aqui proposta visa a apoiar a avaliação de artefatos produzidos em TCCs ao estabelecer um processo básico para elaboração de questionários de avaliação e análise estatística de resultados. Nesta seção, apresentamos as etapas componentes da realização de nossa abordagem.

Na abordagem, a premissa levantada é a de que a avaliação dos artefatos produzidos seja parte integrante do TCC. Neste diapasão, conforme apresentado na Figura 1, a abordagem é constituída das seguintes etapas: (i) especificação; (ii) composição; (iii) execução; (iv) análise. Na realização das etapas de nossa abordagem, está prevista a realização de vários ciclos. Entretanto, em virtude da duração típica de realização de um TCC, não é comum que eles ocorram mais de uma vez.

A etapa de especificação está fortemente relacionada ao que deve ser medido e ao estabelecimento do escopo do TCC. Na especificação, o objetivo é identificar os fatores, os parâmetros e as variáveis de resposta relevantes à avaliação. O detalhamento acerca de como esta etapa deve ser realizada foge ao escopo deste trabalho, uma vez que é uma questão de estratégia e planejamento realizado entre os alunos e os seus orientadores e sua forma de trabalho. De todo modo, é adequado que a etapa de especificação do experimento seja planejada logo no início do TCC.

Desta forma, a avaliação se torna um dos elementos vitais para auxiliar a definição do escopo dos artefatos a serem produzidos [Wainer 2007, Wazlawick 2009, Zobel 2004]. As demais etapas de nossa abordagem são detalhadas ao longo das subseções a seguir.

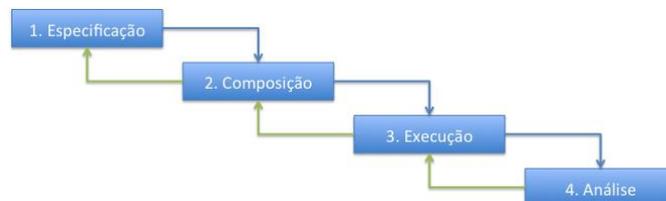


Figura 1. Processo de avaliação do artefato.

3.1 Composição

A composição compreende a definição mais precisa do que deve ser medido e de como deve ser conduzida a avaliação. Quando as variáveis de resposta são quantitativas provenientes de medições diretas, a condução da avaliação costuma ser mais fácil, pois depende basicamente dos alunos que estão elaborando o TCC. Em linhas gerais, a composição constitui-se no estabelecimento dos critérios para seleção de amostras e das replicações de execução da avaliação e coleta das métricas quantitativas que podem ser diretamente medidas com o uso do artefato [Alexopoulos 2007, Carson 2004, Law 2007].

Como muitos TCCs são associados ao desenvolvimento de artefatos que são programas, esses demandam uma avaliação qualitativa, na qual a opinião dos sujeitos acerca da unidade experimental é importante. Assim, entendemos que é necessário adotar uma forma mais precisa e científica para obtenção destes dados. Esta obtenção tem como ponto de partida o estabelecimento de questionários de avaliação de artefatos [Juristo and Moreno 2010]. Estes questionários servem como uma ponte entre os sujeitos e os alunos que desenvolveram os artefatos. Os questionários também

propiciam aos sujeitos a oportunidade de informar sua opinião sobre os artefatos como: o que foi bom, o que foi ruim e o que poderia melhorar.

Para montar um questionário de avaliação, primeiro há que se definir as questões que se pretende abordar. Estas questões devem ser predominantemente concebidas na etapa de especificação e refinadas nesta etapa. Os questionários devem conter um conjunto de perguntas que antecedem a avaliação, denominadas perguntas de pré-execução que visam a qualificar o sujeito. Devem também conter um conjunto de perguntas que são referentes à execução da avaliação. Finalmente, devem também conter um conjunto de perguntas relacionadas à pós-execução.

Como exemplo, considere um TCC que apresenta o desenvolvimento de um sistema de informação (aplicativo). Na região de pré-execução, podem ser apresentadas questões sobre escolaridade, se ele detinha um conjunto de competências, dentre outros itens relevantes para o estudo. Durante a execução da avaliação, o questionário pode solicitar ao usuário, por exemplo, que indique o tempo de uso, se o aplicativo fez o que havia sido proposto e se apresentou defeitos durante o uso. Finalmente, na fase de pós-execução, pode-se, por exemplo, apresentar ao sujeito perguntas comparativas entre o artefato produzido no TCC e outras técnicas conhecidas. Ainda no conjunto de perguntas de pós-execução, é possível também apresentar algumas questões referentes à produção do artefato propriamente dito, como perguntas sobre sua usabilidade ou vinculadas aos testes de hipóteses.

A coleta e gestão de dados por meio de questionários de avaliação pode se tornar uma atividade muito laboriosa. Para mitigar este problema, existem alguns softwares que servem para montar questionários de avaliação. A abordagem proposta faz-se valer de um software gratuito para elaboração de questionários (do inglês, *Forms*) presente no *Google Drive* (GD) [Google 2014]. O GD permite o armazenamento, a criação e a publicação de questionários facilmente na web. A única restrição para uso do GD é que o autor do questionário tenha uma conta de usuário na Google. O GD permite grande variedade de personalização dos questionários produzidos, tanto em seu visual quanto nas perguntas propriamente ditas.

3.2 Execução

Uma vez concluído o planejamento da avaliação, que inclui a especificação e a composição, tem início a etapa de execução. A execução da avaliação compreende a realização de ações para uma correta coleta dos dados. Dentre as ações realizadas, uma das mais importantes está relacionada à seleção das amostras da população de usuários que devem fazer parte da avaliação. As amostras selecionadas devem ser o menos enviesadas possível. Principalmente quando o estudo for comparativo (pareado), e constituído pela execução de atividades envolvendo mais de um artefato, na qual a ordem de contato com os artefatos pode influenciar no estudo.

O número de respostas ao questionário de avaliação também é um fator importante. Isso porque quanto mais respostas o questionário tiver, menor será o viés estatístico correspondente e, conseqüentemente, maior será a validade da análise do estudo. Visto que várias análises estatísticas dependem do estimador da variância, a quantidade mínima de respostas para se considerar o questionário como “suficientemente respondido” é 30 (trinta). Essa quantidade mínima é justificada teoricamente pelo Teorema do Limite Central [Walpole 2012]. Coletado esse número mínimo de respostas, as variações nos resultados tendem a diminuir sensivelmente. No caso da

utilização da ferramenta GD, o início da etapa de execução ocorre com o fornecimento ao sujeito do endereço eletrônico (URL) do questionário correspondente ao artefato avaliado[†].

Se o preenchimento do questionário for realizado em um ambiente controlado (e.g., em um laboratório, com a presença dos elaboradores), durante a condução da avaliação o elaborador do questionário pode auxiliar os sujeitos a responder, desde que não interfira nas respostas selecionadas, o que produziria um viés no resultado. Por outro lado, se o preenchimento for realizado pelos sujeitos por meio da web (i.e., sem interação com o condutor da avaliação), o questionário tem que ser preciso nas suas instruções de preenchimento. Neste caso, é adequado realizar uma execução piloto em um laboratório com dois ou três usuários antes de disponibilizar o questionário à comunidade como um todo. Essa execução piloto tem a finalidade de refinar as instruções de preenchimento, em função de observações e dúvidas geradas durante o teste. Note-se que o preenchimento das respostas fornecidas durante a execução piloto devem ser desconsideradas, ou seja, não podem fazer parte dos dados a serem usados posteriormente para análise.

3.3 Análise

A partir dos dados coletados como resultado da etapa de execução, é possível iniciar a etapa de análise. Os dados coletados (i.e., as respostas dos sujeitos) ficam disponíveis no GD. Na abordagem proposta, a análise é dividida em duas fases: exploratória e de inferência estatística.

A análise exploratória começa diretamente a partir da interface do GD. O GD propicia a exploração inicial dos resultados da pesquisa, questão a questão, por meio de gráficos automaticamente gerados. O tipo de gráfico apresentado no GD é automaticamente elaborado de acordo com o formato previamente proposto para a questão. A interface do GD também possibilita exportar os dados em formato de planilha do Microsoft Excel. O experimentador pode decidir, então, baixar essa planilha para tratar os dados localmente. A nossa abordagem propõe o uso de outra planilha pré-definida em formato Microsoft Excel que possibilita realizar tanto a análise exploratória quanto a análise de inferência estatística.

No que tange à análise exploratória dos dados, a planilha pré-definida oferece o cálculo e apresentação de *Box-Plot*, histogramas e medidas de resumo e dispersão. A Figura 2(a) apresenta a planilha na perspectiva de análise exploratória de dados. Os campos apresentados à esquerda (Entrada) são preenchidos pelo analista, que pode copiá-los da planilha produzida pela GD. A região superior direita apresenta os resultados dos cálculos fornecidos na planilha. A planilha também apresenta o *Box-Plot* de duas amostras (tempo de cadastro e tempo de promoção), possibilitando uma comparação visual das duas amostras. Esta visão geral obtida na fase exploratória, auxilia a postulação de hipóteses.

Uma vez que a etapa exploratória dos dados foi realizada, dá-se início à fase de inferência estatística. A planilha que propomos em nossa abordagem possibilita realizar testes de hipótese com uma ou duas amostras, sejam elas quantitativas ou qualitativas ordinárias. Uma das contribuições deste trabalho é justamente consolidar em uma única planilha a sistemática de elaboração de inferências estatísticas básicas. Esta visão

[†] Um exemplo de questionário é encontrado na página <https://sourceforge.net/p/gpca/wiki/RFIDBook>

consolidada potencializa a tarefa dos alunos, uma vez que eles podem se concentrar na postulação e verificação das hipóteses. Para se realizar a inferência por meio dessa planilha, é necessário preencher três campos e pressionar o botão “*Compute*”, conforme apresentado na Figura 2(b). Os campos que devem ser preenchidos são os contidos na parte superior da imagem (Configuração). O primeiro campo a ser preenchido é o *número de amostras*, que avalia se a análise é em uma ou duas amostras. O segundo campo é o *paramétrica* que serve para indicar se as amostras são paramétricas ou não-paramétricas. Finalmente, o campo *pareado* indica se as amostras possuem dependência, *i.e.*, a coleta de dados de amostras extraídas de populações que são relacionadas uma à outra. Normalmente isso ocorre quando as grandezas vieram do mesmo sujeito. Existem outros dois campos que podem ser preenchidos, mas cujos valores previamente configurados são satisfatórios para a maioria dos casos práticos. São eles: o nível de significância e a diferença de média, cujos valores por omissão (*default*) são, respectivamente, 5% e zero.

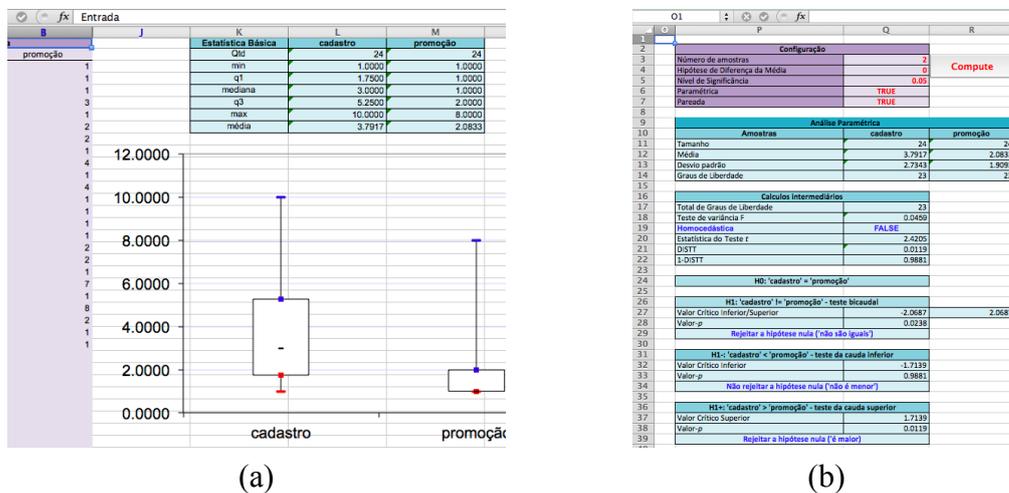


Figura 2. Planilha para exploração de dados (a) e inferência estatística (b)

Os resultados da inferência estatística na planilha são apresentados abaixo da seção intitulada “Configuração”. Para estes resultados, a planilha inicialmente informa se as amostras têm distribuição homocedástica (médias e variâncias semelhantes) ou heterocedástica (médias e variâncias diferentes). No exemplo da Figura 2(b), duas amostras são comparadas (tempo de cadastro e tempo de promoção). Observe que as amostras têm distribuição heterocedástica e está sendo feita uma análise pareada e paramétrica. Neste caso, considera-se verdadeira a hipótese nula (H_0) de que as amostras têm média igual. Se a hipótese alternativa H_1 for que as amostras são diferentes, a hipótese nula é refutada, corroborando com a hipótese alternativa H_1 , *i.e.*, as amostras são diferentes. Se a hipótese alternativa H_{1-} for que o tempo médio para cadastro é menor que o tempo médio de promoção, nota-se que não há evidências que a hipótese nula possa ser refutada. Finalmente, se a hipótese alternativa H_{1+} for que a média da amostra de cadastro é maior que a média de promoção, a hipótese nula é refutada, corroborando com a hipótese alternativa de que o tempo médio de cadastro é maior que o tempo médio de promoção.

Pode-se observar que o resultado alcançado na Figura 2(b) por meio da inferência, intuitivamente podia ser observado no *Box-Plot* apresentado na Figura 2(a). Entretanto, existem diversas outras situações nas quais esta observação visual não é

suficiente. Ademais, pode-se estabelecer que, com *p-value* de 5%, as afirmações obtidas na inferência agregam uma maior credibilidade ao estudo realizado.

4 Aplicação da Abordagem Proposta

A abordagem aqui proposta foi aplicada durante a execução de dois TCCs. O primeiro foi intitulado *RFIDBook*, um sistema de bonificação baseado em tecnologia RFID integrado ao Facebook. O segundo, intitulado *Ame*, é um jogo de cartas para computador baseado no baralho japonês denominado *Hanafuda*.

No uso da abordagem no *RFIDBook*, a etapa de especificação do questionário auxiliou bastante a direcionar os objetivos do aplicativo. Após a execução, foi feita análise exploratória. A composição produziu ao todo 27 perguntas, sendo três qualitativas do sujeito, três quantitativas, 15 qualitativas ordinárias e as demais textuais. A execução da avaliação teve a presença de 28 sujeitos (alunos voluntários). O experimento de avaliação durou quatro horas e dez minutos. Foi feito um pré-processamento para tratar as variáveis. Ao todo foram formuladas oito hipóteses, sendo que quatro eram de apenas uma amostra e quatro eram de duas amostras.

No uso da abordagem no *Hanafuda* a especificação foi natural, uma vez que o projeto já possuía uma proposta de avaliação. A composição produziu ao todo 27 perguntas, sendo duas privadas (contato), dez quantitativas, dez qualitativas ordinárias e as demais textuais. A execução da avaliação teve a presença de 31 sujeitos (alunos voluntários). O experimento de avaliação durou cinco horas e trinta e um minutos. Após a execução foi feita análise exploratória. Foi feito um leve pré-processamento para tratar as variáveis. Ao todo foram formuladas 11 hipóteses, sendo que quatro eram apenas de uma amostra e sete eram de duas amostras.

Nos dois TCCs usados como estudos de caso, observamos que a realização pelos alunos das etapas propostas em nossa abordagem permitiu a eles realizarem rapidamente os experimentos de avaliação necessários. Ao mesmo tempo, as análises estatísticas resultantes valorizaram os artefatos produzidos. Atualmente, a abordagem vem sendo aplicada em mais outros cinco TCCs.

5 Conclusão

A abordagem de avaliação aqui proposta foi utilizada para avaliação de artefatos gerados em dois TCCs. Nessas duas situações, essa realização ocorreu sem comprometer o andamento das demais atividades do TCC. Além disso, a realização da etapa de composição dos questionários, conforme preconizado em nossa abordagem, e da utilização da planilha proposta para análise estatística dos resultados possibilitou aos alunos autores desses TCCs se concentrarem no desenvolvimento e na elaboração do texto do projeto. Acreditamos que isso ocorreu pelo fato das ferramentas adotadas na abordagem serem de uso simples. Os alunos, uma vez treinados, conseguiram realizar a análise exploratória dos dados e a condução das análises estatísticas necessárias. Esse resultado propicia aos egressos dos cursos o acesso a ferramentas práticas de avaliação de artefatos, aos quais muitos terão contato em sua futura vida profissional.

Outro aspecto importante de ressaltar é que os alunos que desenvolveram os TCCs usados como estudos de caso e os respectivos questionários de avaliação não foram os mesmos que conduziram o experimento. Isso foi feito com o propósito de diminuir algum eventual viés de afetividade. De fato, em algumas avaliações, as ferramentas

elaboradas nesses TCCs sofreram diversas críticas. Tal fato poderia não ter ocorrido se o próprio desenvolvedor tivesse aplicado a avaliação.

A aplicação preliminar da abordagem de avaliação a artefatos de dois TCCs, *RFIDBook* e *Hanafuda*, mostrou resultados que indicam um incremento da relação do envolvimento dos alunos no desenvolvimento dos artefatos, pois suscitou neles um sentimento de comprometimento com os resultados e os motivou a desenvolver artefatos com melhores características de usabilidade e utilidade. Além disso, a adoção da abordagem resultou na confecção e apresentação de TCCs mais adequada do ponto de vista científico, uma vez que estes continham uma análise estatística das avaliações realizadas sobre seus respectivos artefatos.

Referências

- Alexopoulos, C., (2007) "Statistical analysis of simulation output: State of the art", In: Simulation Conference, 2007 Winter, p. 150–161
- Boyle, E. A., MacArthur, E. W., Connolly, T. M., Hainey, T., Manea, M., Kärki, A., van Rosmalen, P., (2014) "A narrative literature review of games, animations and simulations to teach research methods and statistics", *Computers & Education*, v. 74 (May.), p. 1–14.
- Carson, J. S., (2004) "Introduction to modeling and simulation", In: Proceedings of the 2004 Winter Simulation Conference, 2004, p. 16
- Google, (2014) "Google Drive", <https://drive.google.com>.
- Juristo, N., Moreno, A. M., (2010), "Basics of Software Engineering Experimentation", 1st ed. Springer Publishing Company, Incorporated.
- Law, A. M., (2007) "Statistical analysis of simulation output data: the practical state of the art", In: Proceedings of the 39th conference on Winter simulation, p. 77–83, Washington D.C.
- Sjoeberg, D. I. K., Hannay, J. E., Hansen, O., Kampenes, V. B., Karahasanovic, A., Liborg, N.-K., Rekdal, A. C., (2005) "A survey of controlled experiments in software engineering", *IEEE Transactions on Software Engineering*, v. 31, n. 9, p. 733–753.
- Tullis, T., Albert, B., (2008) "Self-Reported Metrics", *Measuring the User Experience*, Elsevier, p. 123–166.
- Wainer, J., (2007), "Métodos de pesquisa quantitativa e qualitativa para a Ciência da Computação", *Atualizações em informática*, p. 221–262.
- Walpole, R. E., (2012) "Probability & statistics for engineers & scientists", Boston, Prentice Hall.
- Wazlawick, R., (2009) "Metodologia de Pesquisa em Ciência da Computação", Elsevier Brasil.
- Wohlin, C., Runeson, Per, Höst, Martin, Ohlsson, Magnus C, Regnell, B., Wesslén, Anders, (2012) "Experimentation in Software Engineering", New York, Springer.
- Zobel, J., (2004) "Writing for Computer Science", 2nd ed. Springer.