

Preliminary Guidelines for Empirical Research in Software Engineering

Kitchenham, B.A., Pfleeger, S.L., Pickard, L. M.,
Jones, P.W., Hoaglin, D.C., El-Emam, K., and Rosenberg, J.

Aluna: Erica Rodrigues de Oliveira
Professor: Eduardo Figueiredo

06 de setembro de 2012

Perfil dos autores

- ▶ 4 autores
 - ▶ Pesquisadores em engenharia de software com conhecimento em estatística e ciência da computação
 - ▶ Revisores
 - ▶ Participam de pesquisa empírica
- ▶ 1 autor
 - ▶ Estatístico
 - ▶ Aplica métodos estatísticos para resolução de problemas em Engenharia de Software (EngSoft)
- ▶ 1 autor
 - ▶ Médico estatístico
- ▶ 1 autor
 - ▶ Estatístico com interesse em software e computação
 - ▶ Revisor

▶ 2

Estudo empírico

- ▶ “Estudo empírico em engenharia de software é o **uso científico de dados** quantitativos e qualitativos para entender e melhorar o produto de software e processo de desenvolvimento de software.”

Basili, V.; Elbaum, S. (2006) Better Empirical Science for Software Engineering. In: International Conference on Software Engineering.

▶ 3

Introdução

- ▶ Padrão de pesquisa empírica em engenharia de software é pobre
 - ▶ Esta afirmação não é uma crítica de pesquisadores de software em particular
- ▶ Alguns problemas em medicina
 - ▶ “If researchers have difficulty in a discipline such as medicine, which has a rich history of empirical research, it is hardly surprising that software engineering researchers have problems”
- ▶ Autores se concentraram em diretrizes médicas
 - ▶ Médicos estatísticos têm apontado os baixos padrões de análise estatística em seus artigos
- ▶ Razões para erros estatísticos:
 - ▶ Análise estatística é realizada por pessoas com um entendimento inadequado de métodos estatísticos
 - ▶ Falta de conhecimento estatístico na comunidade de pesquisa

▶ 4

Diretrizes

- ▶ Objetivo
 - ▶ Criar diretrizes para auxiliar pesquisadores a evitar as principais armadilhas em suas atividades de pesquisa e apresentar suas pesquisas corretamente
- ▶ Método
 - ▶ Baseadas em diretrizes médicas
 - ▶ Adequado a estudos empíricos da Engenharia Software

▶ 5

Diretrizes

- ▶ “O que fazer” e “o que não fazer”
 - ▶ Contexto do experimento
 - ▶ Projeto do experimento
 - ▶ Condução do experimento e coleta de dados
 - ▶ Análise
 - ▶ Apresentação dos Resultados
 - ▶ Interpretação dos Resultados

▶ 6

Contexto do Experimento

- ▶ Composto por três elementos:
 - ▶ Informações contextuais sobre as circunstâncias industriais em que um estudo empírico ocorre ou em que uma nova técnica é desenvolvida
 - ▶ Análise das hipóteses da pesquisa e como elas foram obtidas
 - ▶ Informações sobre pesquisas relacionadas
- ▶ Objetivo
 - ▶ Garantir que os objetivos da pesquisa estão corretamente definidos
 - ▶ Garantir que a pesquisa provê detalhes suficientes para outros pesquisadores

▶ 7

Contexto do Experimento – Diretrizes

- ▶ **C1:** *Certificar de especificar, o máximo possível, o contexto industrial. Em particular, definir claramente as entidades, os atributos e medidas que estão captando a informação contextual. Podem ser de dois tipos:*
 - ▶ *Estudo por observação*
 - ▶ Identificar os fatores específicos que podem afetar a generalidade e a utilidade das conclusões (i.e. setor em que os produtos são usados)
 - ▶ Replicação dos resultados em situações específicas
 - ▶ **DESAFIO:** em EngSoft as medidas não são padronizadas (contagem por ponto de função, linha de código)
 - ▶ *Experimentos formais*
 - ▶ Visão geral do ambiente para identificar a melhor técnica (ou conjunto) a ser utilizada

▶ 8

Contexto do Experimento – Diretrizes

- ▶ **C2:** *Se uma hipótese específica está sendo testada, indique-a claramente antes de realizar o estudo, e discuta a teoria a partir do qual ela é obtida, de modo que suas implicações sejam evidentes.*
- ▶ **C3:** *Se a pesquisa é exploratória, especifique de forma clara, e antes da análise dos dados, quais questões a investigação se destina a tratar, e como ela vai lidar com elas.*
 - ▶ Definir as questões de pesquisa com antecedência

▶ 9

Contexto do Experimento – Diretrizes

- ▶ **C4:** Descrever a pesquisa que é semelhante, ou tenha uma influência sobre a pesquisa atual e como o trabalho atual se relaciona com ela.

▶ 10

Projeto do Experimento

- ▶ Descreve os produtos, recursos e processos envolvidos, incluindo:
 - ▶ A população a ser estudada
 - ▶ A lógica e técnica de amostragem da população
 - ▶ O processo de atribuição e gestão dos tratamentos
 - ▶ Os métodos utilizados para reduzir o viés e determinar o tamanho da amostra
- ▶ Objetivo:
 - ▶ Garantir que o projeto é apropriado para os objetivos do estudo.

▶ 11

Projeto do Experimento – Diretrizes

- ▶ **D1:** Identificar a população a partir da qual sujeito e objeto são obtidos
- ▶ **D2:** Definir o processo pelo qual sujeitos e objetos foram selecionados
 - ▶ Justificar o porque da escolha da amostragem
 - ▶ Definir critérios de inclusão e exclusão
- ▶ **D3:** Definir o processo pelo qual sujeitos e objetos são atribuídos aos tratamentos (imparcialidade)
- ▶ **D4:** Restringir-se ao estudo de projetos simples ou, pelo menos, aos projetos que são completamente analisados na literatura. Se você não está replicando o projeto e análise de um estudo anterior, você deve consultar um estatístico para verificar se o seu é o projeto mais eficaz para o que você quer realizar.

▶ 12

Projeto do Experimento – Diretrizes

- ▶ **D5:** Definir a unidade experimental.
- ▶ **D6:** Para os experimentos formais, realize um pré-experimento ou pré-cálculo para identificar ou estimar o tamanho mínimo necessário da amostra.
- ▶ **D7:** Utilizar níveis adequados de blinding.
- ▶ **D8:** Se você não puder evitar a avaliação do seu próprio trabalho, então explicita interesses pessoais (incluindo suas fontes de apoio), e relate o que você tem feito para minimizar o viés.

▶ 13

Projeto do Experimento – Diretrizes

- ▶ **D9:** Evite o uso de controles, a não ser que se tenha certeza que a situação de controle pode ser claramente definida
 - ▶ Exemplo: pedir um grupo para usar uma técnica de projeto novo e outra para não usar nenhuma técnica de projeto
- ▶ **D10:** Definir todos os tratamentos (intervenções)
- ▶ **D11:** Justificar a escolha das medidas de resultados em termos de sua relevância para os objetivos do estudo empírico

▶ 14

Condução do Experimento e Coleta de Dados

- ▶ A condução de um experimento envolve a coleta das medidas dos resultados experimentais.
 - ▶ Problema: experimentos de software não são padronizados
- ▶ **Objetivo**
 - ▶ Garantir que o processo tenha sido bem definido para que o experimento possa ser replicado.

▶ 15

Diretrizes para Coleta de Dados

- ▶ **DC1:** Definir totalmente todas as medidas de software, incluindo a entidade, o atributo, a unidade e as regras de contagem.
 - ▶ Coleta de dados problemática pela falta de padronização
 - ▶ Dificuldade de replicar estudos
- ▶ **DC2:** Para medidas subjetivas, apresente uma medida de concordância entre observadores.
- ▶ **DC3:** Descrever qualquer método de controle de qualidade usado para garantir a plenitude e precisão dos dados coletados.

▶ 16

Diretrizes para Coleta de Dados

- ▶ **DC4:** Para surveys, monitore e relate as taxas de resposta, e discuta a representatividade das respostas e o impacto das não respostas.
- ▶ **DC5:** Para estudos de observação e experimentos, registre dados sobre indivíduos que abandonam os estudos.
- ▶ **DC6:** Para estudos de observação e experimentos, registre dados sobre outras medidas de desempenho que podem ser afetadas de maneira adversa pelo tratamento, mesmo que elas não sejam o foco principal do estudo.
 - ▶ Medicina: registro de dados de efeitos adversos de drogas
 - ▶ Software: produtividade x taxas de defeito

▶ 17

Análise

- ▶ Há duas abordagens para analisar os resultados dos experimentos:
 - ▶ Análise clássica
 - ▶ Análise bayesiana
 - ▶ Utiliza informações/estudos anteriores
 - ▶ Não é muito usada em estudos de engenharia de software
- ▶ **Outra questão**
 - ▶ análise paramétrica: variáveis conhecidas
 - ▶ não paramétrica: variáveis não conhecidas
- ▶ **Objetivo:**
 - ▶ Garantir que os resultados experimentais obtidos são analisados corretamente.

▶ 18

Análise – Diretrizes

- ▶ **A1:** Especificar os procedimentos utilizados para controlar múltiplos testes.
- ▶ **A2:** Considerar o uso de análise blind.
- ▶ **A3:** Realizar análise de sensibilidade.
 - ▶ Identificar e tratar outliers
- ▶ **A4:** Garantir que os dados não violam as suposições de testes utilizados neles
 - ▶ Transformar os dados (teste t – logaritmo)
- ▶ **A5:** Aplicar controle de qualidade apropriado para verificar os resultados
 - ▶ “It’s not an opportunity to discard data or change values to favor your hypothesis.”

▶ 19

Apresentação dos resultados

- ▶ O leitor de um estudo deve ser capaz de compreender a razão do estudo, o projeto, a análise, os resultados e o significado desse estudo
- ▶ Também deve ser capaz de reproduzir ou replicar esse estudo
- ▶ Importante: detalhar os procedimentos utilizados

▶ 20

Apresentação dos resultados – Diretrizes

- ▶ **P1:** Descrever ou citar uma referência para todos os procedimentos estatísticos utilizados.
- ▶ **P2:** Relatar o pacote estatístico utilizado.
- ▶ **P3:** Apresentar resultados quantitativos, assim como níveis de significância. Resultados quantitativos devem mostrar a magnitude dos efeitos e os limites de confiança.
 - ▶ Exemplo: Relatar o valor da estatística t ; Para a regressão, relatar a equação de regressão; Para a regressão relatar o coeficiente de determinação

▶ 21

Apresentação dos resultados – Diretrizes

- ▶ **P4:** Apresentar os dados brutos quando possível. Confirmar que eles estão disponíveis para revisões confidenciais pelos revisores e auditores independentes.
- ▶ **P5:** Fornecer estatística descritiva apropriada.
 - ▶ Relatar o número de observações
 - ▶ Relatar todos os números com o grau adequado de precisão.
 - ▶ Apresentar o numerador e o denominador de porcentagens.
 - ▶ Com números pequenos, apresentar valores e não porcentagens.
 - ▶ Apresentar as medidas apropriadas de tendência central e dispersão quando resumir os dados contínuos.

▶ 22

Apresentação dos resultados – Diretrizes

- ▶ **P6:** Fazer o uso adequado dos gráficos
 - ▶ Erros comuns:
 - ▶ Representação unidimensional de dados em duas ou mais dimensões
 - ▶ Utilização de gráficos de pizza
 - ▶ Escolha inadequada da escala
 - ▶ Omissão de pontos distantes em gráficos de dispersão
 - ▶ Omitir dados em gráficos de dispersão, quando muitos pontos de dados se sobrepõem.

▶ 23

Interpretação dos Resultados

- ▶ Todas as conclusões devem ser obtidas diretamente dos resultados
 - ▶ Não se deve introduzir um novo material na seção de conclusões
- ▶ Qualificar os resultados apropriadamente

▶ 24

Interpretação dos Resultados – Diretrizes

- ▶ **I1:** Definir a população na qual modelos estatísticos inferenciais e preditivos se aplicam
 - ▶ Diretamente relacionado a D1 (identificar população)
- ▶ **I2:** Diferenciar entre significância estatística e importância prática
- ▶ **I3:** Definir o tipo de estudo
 - ▶ Exemplos: observação, estudo em ambiente controlado, etc.
- ▶ **I4:** Especificar as limitações do estudo
 - ▶ Validade interna e externa

▶ 25

Resumindo

- ▶ Certifique-se de que o contexto é bem definido
- ▶ Descreva a relação de trabalhos anteriores
- ▶ Certifique-se de que o processo é repetível
- ▶ Verifique se o produto é reproduzível
- ▶ Documente o seu trabalho corretamente, se necessário consulte estatísticos
- ▶ Descreva e discuta a estabilidade e validade dos dados
- ▶ Conheça o seu experimento
- ▶ Utilize as métricas corretas
- ▶ Sempre que possível, apresente os dados brutos
- ▶ Use gráficos adequadamente

▶ 26

Conclusões

- ▶ Diretrizes apresentadas:
 - ▶ Melhorar a qualidade da execução e da avaliação da pesquisa empírica
- ▶ Importância:
 - ▶ Os pesquisadores de software geralmente cometem erros estatísticos
 - ▶ Pesquisadores seniores estão pressionando por mais pesquisas empíricas para apoiar a engenharia de software

▶ 27

Conclusões

- ▶ Pesquisadores podem melhorar suas pesquisas
- ▶ Revisores e leitores podem usar as diretrizes para avaliar a qualidade da pesquisa
- ▶ Algumas diretrizes podem ter implicações éticas e/ou metodológicas
- ▶ As diretrizes não são suficientes por si só

▶ 28