

Testes de Usabilidade

INF1403 – Introdução à Interação Humano-Computador

Prof. Alberto Raposo

abraposo@inf.puc-rio.br

sala 413 RDC

como avaliar > tipos de métodos

investigação

entrevistas

Questionários

Grupos de foco

diários

inspeção

avaliação heurística

percurso cognitivo

inspeção semiótica
(MIS)

observação

direta

testes de usabilidade

avaliação de
comunicabilidade
(MAC)

prototipação em papel

Wizard of Oz

indireta

análise de logs

Engenharia de Usabilidade



- Objetivo é melhorar o sistema a cada versão.
- Nível atual de desempenho.
- Nível mínimo aceitável de desempenho.
- Nível desejado de desempenho.

Teste de Usabilidade

- Avaliação Empírica
 - **Empírico** é um fato que se apoia somente em experiências vividas, na observação de coisas, e não em teorias e métodos científicos
- Objetiva detectar problemas de usabilidade por meio da observação do usuário interagindo com o sistema.

Testes de Usabilidade

- Em geral a expressão é usada para denominar testes de observação de usuários com:
 - protótipos de **baixa fidelidade** (representações provisórias e tentativas de partes da tecnologia sendo testada); e
 - protótipos de **alta fidelidade** (representações muito próximas da versão final da tecnologia sendo testada).
- As observações podem ser feitas:
 - em laboratório (ambiente artificial preparado pelo avaliador)
 - em campo (ambiente natural do usuário)
 - por captura remota de dados (através de tecnologias avançadas de instrumentação e registro de dados)
 - uma combinação das opções acima

Teste de Usabilidade



Envolve gravação do desempenho de usuários típicos, realizando tarefas típicas.

Ambiente controlado.

Usuários são observados e cronometrados.

Dados são gravados em vídeo e apertado de teclas são logados.

Os dados são usados para calcular os tempos gastos pelos usuários nas tarefas, bem como para identificar e explicar erros.

Satisfação do usuário é avaliada por meio de questionários e entrevistas.

Estudos de campo podem ser usados para prover entendimento contextual.



testes de usabilidade tradicionais

dados quantitativos

taxa de sucesso

número de participantes que concluiu cada tarefa

número de erros cometidos, por tipo de erro, por tarefa, por participante

número de vezes em que cada participante consultou ajuda, por tarefa

desempenho

tempo para realização de cada tarefa, por participante

número de itens de menu que o participante acessou antes de encontrar o “correto”

...

dados qualitativos

pontos positivos e negativos

expectativas e quebra de expectativas

...

seis etapas para a condução de um teste

1. desenvolver o plano de teste
2. selecionar e recrutar participantes
3. preparar o material do teste
4. conduzir o teste
5. *debrief* o participante
6. transformar dados em resultados e recomendações

1. plano de teste

2. recrutamento; 3. preparação; 4. condução; 5. *debriefing*; 6. relato

objetivos (gerais) do teste

perguntas que a avaliação visa responder

perfil de usuário

procedimento de teste

ambiente / equipamento de teste

tarefas a serem observadas

dados que serão coletados

procedimento de análise dos dados

conteúdo e apresentação do relatório

1. plano de teste

2. recrutamento; 3. preparação; 4. condução; 5. *debriefing*; 6. relato

objetivos (gerais) do teste

perguntas que a avaliação visa responder

perfil de usuário

procedimento de teste

ambiente / equipamento de teste

tarefas a serem observadas

dados que serão coletados

procedimento de análise dos dados

conteúdo e apresentação do relatório

descobrir se/quantos usuários conseguem realizar as tarefas, em quanto tempo e cometendo quantos erros

verificar se a documentação compensa falhas na interface

avaliar problemas relatados ao suporte

1. plano de teste

2. recrutamento; 3. preparação; 4. condução; 5. *debriefing*; 6. relato

objetivos (gerais) do teste

perguntas que a avaliação visa responder

perfil de usuário

procedimento de teste

ambiente / equipamento de teste

tarefas a serem observadas

dados que serão coletados

procedimento de análise dos dados

conteúdo e apresentação do relatório

Os usuários **conseguem transitar** entre os principais módulos do sistema?

O **tempo de resposta** causa erros ou frustração?

As telas refletem o **modelo conceitual** do usuário?

Que **obstáculos** impedem que a tarefa X seja concluída?

Esta nova versão é **mais fácil** de usar do que a anterior?

Há **equilíbrio** entre fácil de usar e fácil de aprender?

Quanto tempo o usuário leva para:

- encontrar cada produto?
- finalizar a compra, após inseridos todos os produtos no carrinho?
- refazer uma compra realizada anteriormente, com exceção de alguns produtos que foram incluídos por causa de um almoço de família?

1. plano de teste

2. recrutamento; 3. preparação; 4. condução; 5. *debriefing*; 6. relato

objetivos (gerais) do teste

perguntas que a avaliação visa responder

perfil de usuário.....

procedimento de teste

ambiente / equipamento de teste

tarefas a serem observadas

dados que serão coletados

procedimento de análise dos dados

conteúdo e apresentação do relatório

características do usuário

idade, sexo, atitude frente a computadores, estilo de aprendizado (aprender lendo vs. fazendo vs. observando outros)

experiência geral

com computadores, sistema operacional, Internet, domínio da aplicação, aplicações semelhantes, a aplicação sendo testada tempo de uso, frequência de uso, tipos de tarefas realizadas

formação

nível de instrução, área de formação

atividades

função profissional, experiência nessa função, treinamento, rotatividade

atitude

com relação a: domínio da aplicação, aplicações semelhantes, a aplicação sendo testada

e.g., profissionais liberais “conectados”, sem tempo de ir ao supermercado e que costumam comprar os mesmos produtos toda semana

Obs.: usuário ≠ cliente ≠ desenvolvedor

1. plano de teste

2. recrutamento; 3. preparação; 4. condução; 5. *debriefing*; 6. relato

objetivos (gerais) do teste

perguntas que a avaliação visa responder

perfil de usuário

procedimento de teste

ambiente / equipamento de teste

tarefas a serem observadas

dados que serão coletados

procedimento de análise dos dados

conteúdo e apresentação do relatório

recepção do participante e explicações sobre o teste

questionário pré-teste

(para obter o perfil do participante)

termo de consentimento livre e esclarecido (TCLE)

orientação

script de introdução e orientação para o teste; objetivo; garantia de anonimato; forma de observação (registro de áudio/vídeo)

teste (de desempenho / preferências)

cenários de tarefas
o que será medido

debriefing do participante

questionário e entrevista pós-teste
comentários gerais sobre o produto / preferências

1. plano de teste

2. recrutamento; 3. preparação; 4. condução; 5. *debriefing*; 6. relato

objetivos (gerais) do teste

perguntas que a avaliação visa responder

perfil de usuário

procedimento de teste

ambiente / equipamento de teste.....

tarefas a serem observadas

dados que serão coletados

procedimento de análise dos dados

conteúdo e apresentação do relatório

hardware

PC; tablet; smartphone

webcam; filmadora

microfone; gravador de áudio;

eyetracker

software

objeto do teste

sem nenhum dado de teste

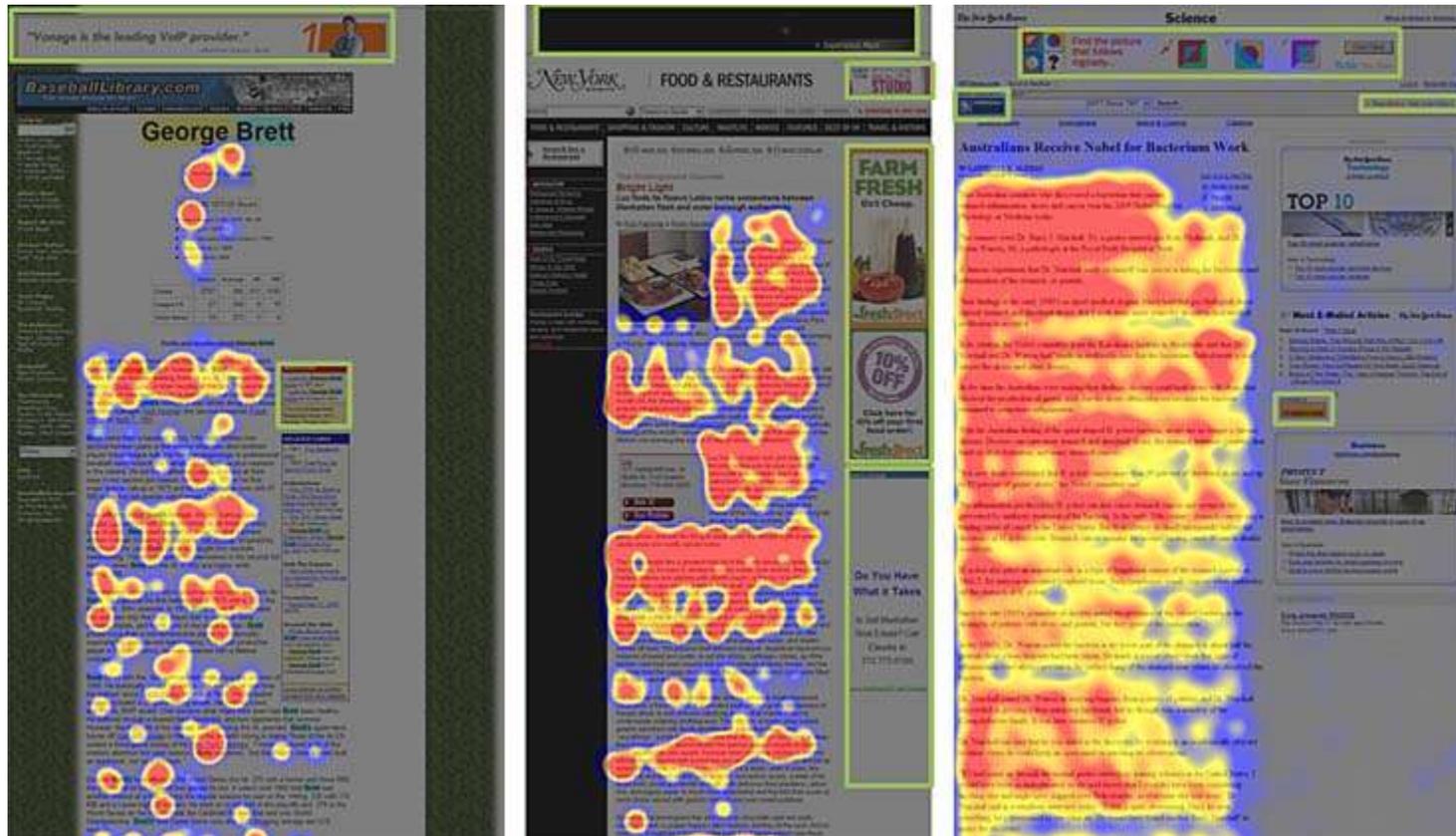
anterior que possa comprometer

o atual (*e.g.*, após limpar “cache”

e restaurar a base de dados)

para captura de tela

Exemplo de dados coletados com rastreamento ocular



Mapas de estudos de eyetracking: as áreas onde os usuários mais olhavam estão em vermelho; As áreas amarelas indicam menos visões, seguidas pelas áreas azuis. As áreas cinzentas não atraíram nenhuma fixação. Caixas verdes foram desenhadas em cima das imagens após o estudo para destacar os anúncios.

<https://www.nngroup.com/articles/banner-blindness-old-and-new-findings/>

1. plano de teste

2. recrutamento; 3. preparação; 4. condução; 5. *debriefing*; 6. relato

objetivos (gerais) do teste

perguntas que a avaliação visa responder

perfil de usuário

procedimento de teste

ambiente / equipamento de teste

tarefas a serem observadas

dados que serão coletados

procedimento de análise dos dados

conteúdo e apresentação do relatório

tarefas

- quais: priorizadas por frequência, criticidade, vulnerabilidade, disponibilidade
- em quais módulos/versões do sistema
- para cada tarefa: pré-requisitos, critérios de sucesso, duração máx. de execução

cuidados

- divisões equilibradas (vs. especialistas no mesmo grupo)
- independência entre os grupos
- descrição das tarefas (o que fazer, mas não como; evitando menções a elementos de

Grupo	Sistema (X,Y) – Tarefa (A, B)
G1	X-A, Y-B
G2	X-B, Y-A
G3	Y-A, X- B
G4	Y-B, X-A

1. plano de teste

2. recrutamento; 3. preparação; 4. condução; 5. *debriefing*; 6. relato

objetivos (gerais) do teste

perguntas que a avaliação visa responder

perfil de usuário

procedimento de teste

ambiente / equipamento de teste

tarefas a serem observadas

dados que serão coletados.....

procedimento de análise dos dados

conteúdo e apresentação do relatório

medidas de **desempenho**

tempo

- para completar uma tarefa
- para acessar informação no manual / na ajuda online
- lendo uma seção do manual
- conversando com o atendimento ao usuário

número de ...

- tarefas concluídas corretamente / sem ajuda (e %)
- seleções incorretas
- erros de omissão
- seleções de menu incorretas
- ícones incorretos
- chamadas ao *help desk*

1. plano de teste

2. recrutamento; 3. preparação; 4. condução; 5. *debriefing*; 6. relato

objetivos (gerais) do teste

perguntas que a avaliação visa responder

perfil de usuário

procedimento de teste

ambiente / equipamento de teste

tarefas a serem observadas

dados que serão coletados.....

procedimento de análise dos dados

conteúdo e apresentação do relatório

medidas de **preferências**

- utilidade do produto
- o quão bem o produto satisfaz as expectativas
- adequação das funções do produto às tarefas dos usuários
- facilidade de uso
- facilidade de aprendizado
- facilidade de instalação e configuração
- utilidade da ajuda
- respostas do atendimento ao usuário
- protótipo A vs. protótipo B
- produto A vs. competidor B
- modelo conceitual de A vs. de B

coletadas através de

- entrevista (semiestruturada) ou
- questionário (com escala de Likert de 7 pontos + justificativa como pergunta aberta)

Alguns tipos de dados coletados

- Tempo pra completar uma tarefa.
- Tempo pra completar uma tarefa após determinado tempo sem usar o produto.
- Número e tipos de erros por tarefa.
- Número de erros por unidade de tempo.
- Número de interações necessárias para help online ou manuais.
- Número de usuários cometendo determinado erro.
- Número de usuários completando a tarefa com sucesso.

1. plano de teste

2. recrutamento; 3. preparação; 4. condução; 5. *debriefing*; 6. relato

objetivos (gerais) do teste

perguntas que a avaliação visa responder

perfil de usuário

procedimento de teste

ambiente / equipamento de teste

tarefas a serem observadas

dados que serão coletados

procedimento de análise dos dados

conteúdo e apresentação do relatório

estatísticas

triangulação entre dados observados, depoimentos espontâneos e dados da entrevista

1. plano de teste

2. recrutamento; 3. preparação; 4. condução; 5. *debriefing*; 6. relato

objetivos (gerais) do teste

perguntas que a avaliação visa responder

perfil de usuário

procedimento de teste

ambiente / equipamento de teste

tarefas a serem observadas

dados que serão coletados

procedimento de análise dos dados

conteúdo e apresentação do relatório

estrutura do documento

- planejamento
- execução
- descrição e análise dos resultados
- recomendações

2. recrutamento de participantes

1. plano; _____; 3. preparação; 4. condução; 5. *debriefing*; 6. relato
perfil

correspondência com um dos perfis de usuário-alvo
questionário para o recrutamento

disponibilidade

tempo

localização geográfica

Não subestimem o recrutamento!

Quantos usuários são necessários para testes de usabilidade?



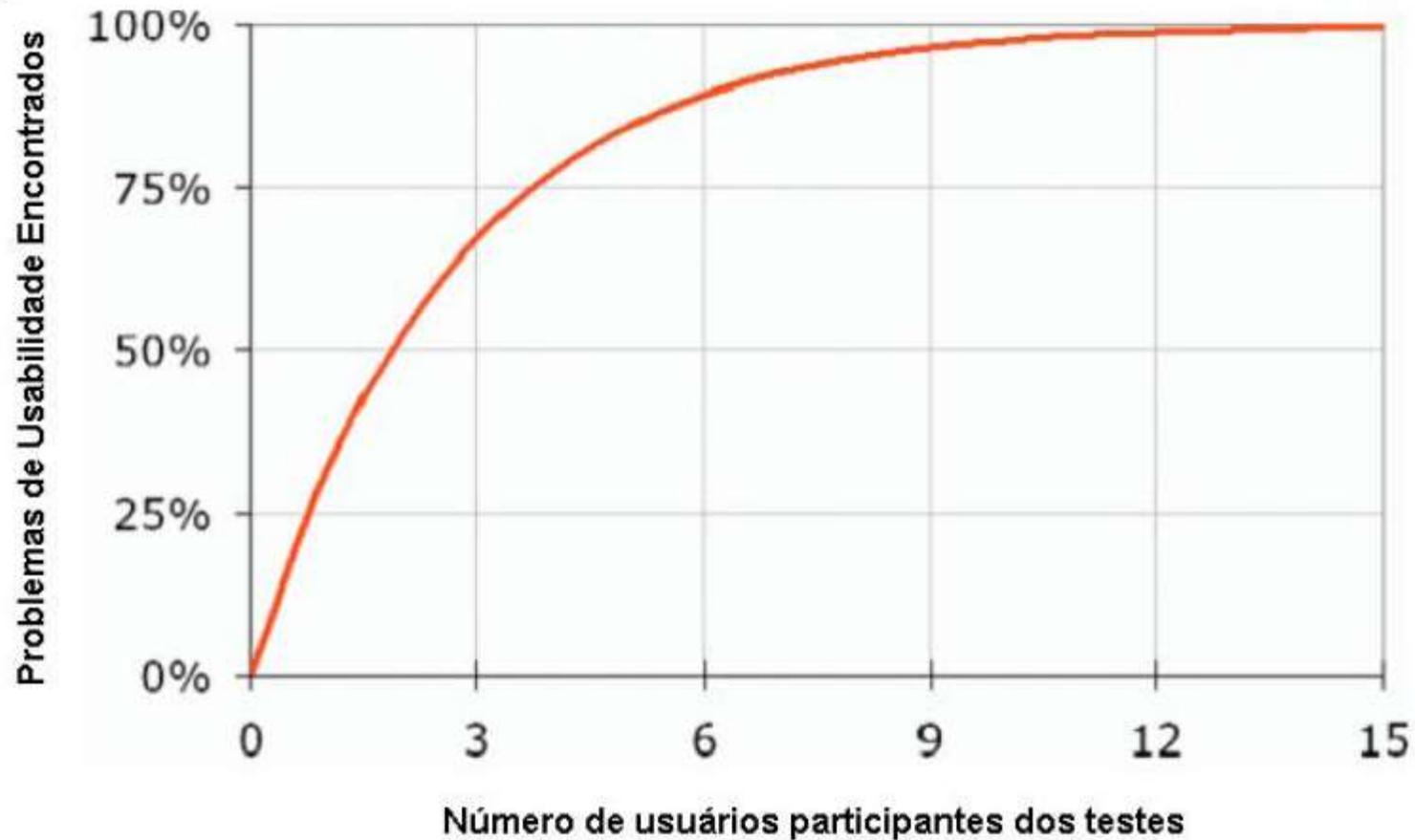
O número é uma questão de ordem prática.
Depende de:

- Prazo para os testes;
- Disponibilidade de participantes;
- Custo de execução dos testes.

Tipicamente 5-10 participantes.

Alguns especialistas argumentam que teste deve continuar até que nenhum novo insight seja obtido.

Custo x benefício para execução de testes



3. preparação do material para o teste

1. plano; 2. recrutamento; _____; 4. condução; 5. *debriefing*; 6. relato

roteiro (*script*) de orientação

questionário pré-teste (perfil, expectativas)

instrumentos de coleta de dados (hardware e software)

termo de consentimento livre e esclarecido (TCLE)

(e, em alguns casos, NDA – *non-disclosure agreement*)

materiais de treinamento necessários

descrição das tarefas

(sem detalhes que interfiram no resultado do teste)

questionário pós-teste

roteiro de tópicos para entrevista pós-teste

4. condução do teste

1. plano; 2. recrutamento; 3. preparação; _____; 5. *debriefing*; 6. relato

trate cada novo participante como um indivíduo único

- e não apenas “mais um”

mantenha um ambiente tranquilo, relaxado

- deixe claro para o participante que não há resposta certa ou errada

monitore a sessão de modo imparcial

- esteja consciente dos efeitos da sua voz e linguagem corporal
- evite demonstrar surpresa ou emoção, mesmo quando o participante erra ou faz algo inesperado

evite ajudar o participante

- não “resgate” o participante quando ele estiver confuso ou perdido

se você cometer um erro, continue como se fosse normal

- e.g. dar dicas ou revelar informações que induzam certos comportamentos

se apropriado, utilize a técnica de “*think aloud*”

- anote os depoimentos espontâneos do participantes

Teste Piloto

- Testes piloto devem ser conduzidos para assegurar-se que todo o experimento esteja pronto. Podem ser conduzidos com a própria equipe de desenvolvimento.
- Testes pilotos evitam “queimar” a equipe de avaliação perante os participantes com testes mal preparados - bons participantes são difíceis de serem conseguidos.

5. *debriefing* (conversa/questionamento pós-sessão)

1. plano; 2. recrutamento; 3. preparação; 4. condução; _____; 6. relato
questionário pós-teste

enquanto o participante preenche o questionário pós-teste:

organize suas ideias com base no roteiro de entrevista e no que foi observado durante a sessão

dê uma olhada no questionário pós-teste respondido

marque pontos para explorar na entrevista

entrevista

1. deixe o participante dizer o que pensa, livremente

“O que você achou?”

2. questões mais gerais

3. questões específicas

revise pontos que você marcou no questionário pós-teste

se concentre em entender problemas e dificuldades

e não em resolver problemas

6. relato > compilação dos dados coletados

1. plano; 2. recrutamento; 3. preparação; 4. condução; 5. *debriefing*;

–ao final de cada sessão de teste:

 passe dados escritos para meio digital, transcreva trechos interessantes da fala dos participantes

sumarize dados de desempenho

 tempo de realização de cada tarefa

 médio, mín/máx, desvio padrão

 acurácia: porcentagem de participantes que concluiu a tarefa com sucesso

 dentro do tempo estimado; fora do tempo estimado; sem buscar ajuda ou tendo buscado ajuda

sumarize dados de preferência

 contagem das respostas de múltipla escolha

 listas de comentários nas perguntas abertas

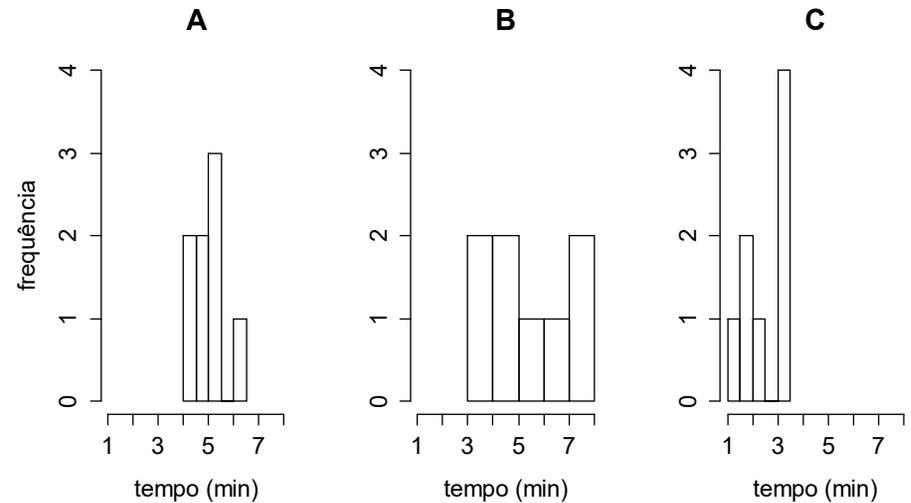
 trechos de transcrição da entrevista pós-teste

6. relato > compilação dos dados coletados

1. plano; 2. recrutamento; 3. preparação; 4. condução; 5. *debriefing*

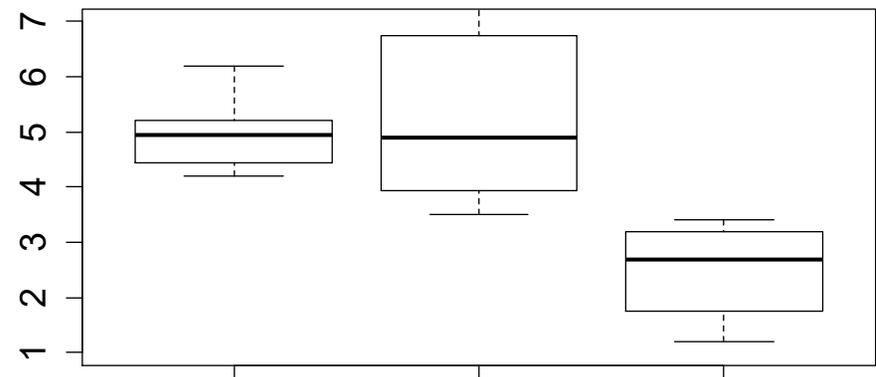
Tarefa T1, tempo (min), sistemas A, B e C

grupo participante	A	B	C
g1 p1	4.7	4.2	3.4
g1 p2	4.2	5.3	3.2
g1 p3	5.2	3.5	2.3
g1 p4	4.6	3.7	3.1
g2 p5	6.2	7.4	3.2
g2 p6	4.3	4.5	1.2
g2 p7	5.2	6.2	1.7
g2 p8	5.2	7.3	1.8



	A	B	C
Min.:	4.200	3.500	1.200
1st Qu.:	4.525	4.075	1.775
Median:	4.950	4.900	2.700
Mean:	4.950	5.263	2.487
3rd Qu.	5.200	6.475	3.200
Max.	6.200	7.400	3.400

Tempo (min) para realização de T1



Testes t indicaram que o desempenho com o sistema C é significativamente melhor que os desempenhos com o sistema A e com o sistema B ($p < 0.05/3$, com a correção de Bonferroni).

cuidado com conclusões de base “estatística”



6. relato > análise dos dados coletados

1. plano; 2. recrutamento; 3. preparação; 4. condução; 5. *debriefing*

conforme grupo de participantes ou versões do sistema

analise diferenças entre grupos ou versões de produto

tarefas que não satisfizeram os critérios desejados

erros e dificuldades dos usuários

faça uma análise da origem dos erros

priorize os problemas por gravidade e probabilidade de ocorrência

elabore recomendações

indicando pontos que precisam ser pesquisados mais profundamente

6. relato > conteúdo e apresentação

1. plano; 2. recrutamento; 3. preparação; 4. condução; 5. *debriefing*

plano de teste

resultados sumarizados

estatísticas básicas dos dados quantitativos

[se distribuição for próxima da normal] média, desvio padrão

[se distribuição for diferente da normal] mín, Q1, mediana, Q3, máx

[se houver hipóteses] significância dos resultados

tabelas e gráficos

resultados

recomendações e discussão



testes de usabilidade