

Experimentos: Exemplos, Ferramentas e Boas Práticas

Prof.: Leonardo Tórtoro Pereira

Como saber o que devo coletar?

Como saber o que devo coletar?

- Seu objetivo, hipóteses e metodologia de pesquisa devem dar uma ideia de quais dados são necessários para comprovar ou descartar a hipótese e, com isso, tentar alcançar o objetivo da pesquisa
- Porém, qual o melhor meio para obter tais dados?
- Como evitar viés?
- Como tratar os dados?
- Como analisá-los?

Obtendo Informações de Usuários

Obtendo Informações de Usuários [1]

- De modo geral, temos 2 meios de obter dados de usuários
 - ◆ Informações que os usuários fornecem explicitamente ao sistema
 - ◆ Informações que o sistema obtém implicitamente

Obtendo Informações de Usuários Explicitamente

Obtendo Informações de Usuários [1]

- As informações explicitamente fornecidas podem vir de diferentes maneiras
 - ◆ **Relatório Pessoal Sobre Características Pessoais Objetivas**
 - ◆ **Auto-avaliação com Respeito a Dimensões Gerais**
 - ◆ **Relatório Pessoal sobre Avaliações Específicas**
 - ◆ **Respostas para itens de testes (provas)**

Relatório Pessoal Sobre Características Pessoais Objetivas

Obtendo Informações de Usuários [1]

- Informações sobre propriedades objetivas do usuário
 - ◆ Idade, profissão, endereço
- Se os dados podem levar à identificação do usuário, é preciso pedir permissão ao conselho de ética
- Muitas vezes, é mais prático (e não afeta o resultado) pedir para usuário escolher em qual grupo pertence ao invés de uma resposta extensa
 - ◆ Ex: grupos de idade ao invés da idade em específico

Obtendo Informações de Usuários [1]

- Alguns usuários podem não querer dar tais dados
 - ◆ Nesses casos, é preciso definir se são necessários ao experimento ou não
 - Se forem, usuário será descartado ao não fornecer
 - Se não forem, dar opção de pular tais dados e coletar o resto

Auto-avaliação com Respeito a Dimensões Gerais

Obtendo Informações de Usuários [1]

- Levanta dados sobre usuário em relação à sua posição em uma dimensão geral particular
 - ◆ Interesse do usuário sobre um tópico
 - ◆ Nível de conhecimento sobre algo
 - ◆ Importância que usuário dá a um critério de avaliação específico
 - ◆ ...

Obtendo Informações de Usuários [1]

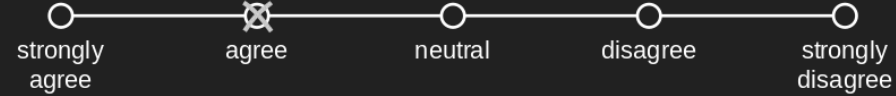
- Dependendo do caso, é possível extrair essas informações implicitamente, caso seu sistema colete tais dados
 - ◆ Porém, pode ser mais rápido perguntar diretamente
 - ◆ Pode ser do interesse da pesquisa confirmar exatamente se seu procedimento metodológico consegue inferir tais dados corretamente

Obtendo Informações de Usuários [1]

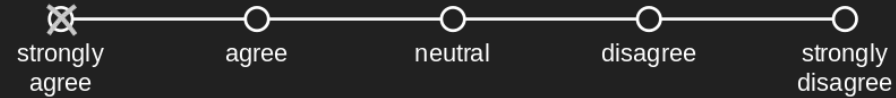
- É importante deixar claro o que todos os pontos das escalas escolhidas significam
- Evitar pressionar para responder algo diferente do que pensa por não ser socialmente aceito
- Anonimidade muitas vezes pode ajudar
- Perguntas claras
- Uso de escala Likert com pontos bem definidos

Website User Survey

1. The website has a user friendly interface.



2. The website is easy to navigate.



3. The website's pages generally have good images.



Fonte: https://en.wikipedia.org/wiki/Likert_scale

Escala Likert [2]

- Quantos pontos usar numa escala Likert?
 - ◆ Depende. O importante é que os pontos deixem claro o que eles descrevem.
 - ◆ Geralmente, 5 pontos é o bastante
 - Menos é difícil de separar bem as opiniões
 - Mais costuma ser difícil de memorizar e pode não ter diferenças claras entre os grupos

Escala Likert [2]

- Quantos pontos usar numa escala Likert?
- ◆ Geralmente, é colocado um número ímpar
 - Mas, se opiniões neutras forem ruins, par é melhor
 - Evita indecisões

Relatório Pessoal sobre Avaliações Específicas

Obtendo Informações de Usuários [1]

- Coleta de informações sobre avaliação explícita do usuário sobre itens específicos
- Itens avaliados podem ser
 - ◆ O que se está tendo uma experiência direta com (jogo, aplicativo, etc.)
 - ◆ Ações que acabou de realizar
 - ◆ Itens que ele precisa julgar na base da descrição
 - ◆ Nome de algo com o qual teve experiência no passado

Obtendo Informações de Usuários [1]

- Costuma ser respondido com um clique do mouse
 - ◆ Porém, pode ser necessária alguma inferência ou recuperação de memória dos usuários
 - ◆ Usuários não gostam de realizar avaliações explícitas que não são diretamente relacionadas à tarefa realizada
 - Idealmente, faça perguntas o mais objetivas possível e de tamanho reduzido

Respostas para itens de testes (provas)

Obtendo Informações de Usuários [1]

- Em alguns casos, queremos administrar testes para saber conhecimentos ou habilidades dos usuários
- Usuários dificilmente irão querer fazer isso por vontade própria
 - ◆ Tente apresentar de maneira divertida
 - ◆ Uma prova “escondida” dentro de alguma funcionalidade mais interessante

Obtendo Informações de Usuários Implicitamente

Informações de Usuários Implícitas [1]

- Ações que ocorrem naturalmente
- Informações salvas previamente
- Índices de baixo nível de estados psicológicos
- Sinais sobre o ambiente atual do usuário

Ações que ocorrem naturalmente

Informações de Usuários Implícitas [1]

- Ações que usuário faz no sistema sem um propósito específico de levantar informações do usuário
 - ◆ Comprar um produto
 - ◆ Rolar por uma página
- As mais significativas tendem a ser específicas ao sistema particular envolvido
 - ◆ Ex: escolha de músicas no Spotify

Informações de Usuários Implícitas [1]

- Em contextos como navegação de hipertextos, pesquisadores já desenvolveram meios para analisar as ações de navegação de um usuário para inferir os interesses do usuário e/ou propor atalhos de navegação
- Algumas áreas têm suas próprias pesquisas
 - ◆ Ex: diversas métricas de game design para identificar calma ou ansiedade durante o jogo
 - [Left 4 Dead - The Director](#)

Informações de Usuários Implícitas [1]

- Apesar de não requererem um investimento adicional do usuário, elas são difíceis de interpretar
 - ◆ Se o usuário abriu uma página/video/música não significa que ele consumiu o conteúdo de fato
- Usar esses dados sem o consentimento do usuário pode trazer problemas de privacidade
 - ◆ Informe sempre os usuários de quais dados são coletados em experimentos!

Informações salvas previamente

Informações de Usuários Implícitas [1]

- Às vezes um sistema pode ter dados do usuário que foram armazenadas independente da interação do usuário com o sistema
 - ◆ Se usuário tem relação com a organização que opera o sistema
 - ◆ Fontes públicas como páginas pessoais
 - ◆ Um modelo do usuário feito com outro sistema

Índices de baixo nível de estados
psicológicos

Informações de Usuários Implícitas [1]

- Dados coletados por sensores que refletem o estado psicológico do usuário
 - ◆ Raiva e frustração
 - Útil para diálogos automatizados
 - ◆ Atração em itens particulares
 - Útil para sistemas de recomendação
 - ◆ Estresse e carga cognitiva
 - Dirigir, provas, coisas difíceis no geral

Informações de Usuários Implícitas [1]

- Dispositivos colocados no corpo do usuário que transmitem dados fisiológicos
 - ◆ Sinais de eletromiograma, resposta galvânica da pele, pressão do sangue, padrões de respiração
- Câmeras ou microfones que transmitem informações fisiológicas
 - ◆ Expressões faciais
 - ◆ Dados da voz

Informações de Usuários Implícitas [1]

- É necessário usar técnicas de reconhecimento de padrões para extrair informações significativas dos dados
 - ◆ Aprendizado de máquina
- Geralmente precisa ser combinado com outros dados para maior acurácia

Sinais sobre o ambiente atual do usuário

Informações de Usuários Implícitas [1]

- Dispositivos que recebem sinais explícitos do ambiente em que o usuário está através de transmissores especializados
 - ◆ GPS
- Sensores gerais também podem fornecer dados sobre o ambiente
 - ◆ Câmeras e Microfones
 - ◆ Precisa de algoritmos para interpretar sinais

Dados sobre Programas

Dados sobre programas

- No caso de dados sobre um algoritmo, aplicativo, modelo, arquitetura, etc. é preciso entender quais as métricas mais aceitas pela literatura para medir o que é desejado
- No caso geral, costuma ser interessante medir o tempo de execução e a acurácia do algoritmo
 - ◆ Porém, em alguns casos, não sabemos qual o resultado esperado (ground truth)
 - Exploração visual dos dados pode ser útil

Dados sobre programas

- Dependendo do domínio de aplicação, algumas métricas poderão ser úteis para validar os resultados
 - ◆ Isso precisa ser definido no design do experimento
 - ◆ Entender quais as melhores métricas demanda uma boa revisão da literatura
 - ◆ Ex:
 - Alcance de expressão, densidade, matriz de erro, curva ROC...

Dados sobre programas

- De modo geral, é necessário executar os testes computacionais diversas vezes e analisar a média e desvio padrão dos resultados para concluir algo
- Análise visual pode ajudar a compreender, mas apenas um teste estatístico adequado pode confirmar se hipótese nula é mantida ou descartada
- Para isso, é preciso ter um grande conjunto de dados para cada entrada (ex: 100 ou 1000 execuções para cada uma)

Visualização de Dados

Visualização de Dados

- Escolher o tipo de visualização adequado pro que você deseja representar é essencial!
- Cada tipo de gráfico tem dados que representam melhor
- [Esse site pode ajudar](#)

Visualização de Dados

- Se você não sabe muito de Python ou R, pode ser mais fácil criar gráficos através de programas de planilha
 - ◆ Excel, Libre Office, Google Sheets...
 - ◆ Todos tem uma boa quantidade de gráficos prontos para usar
 - Só precisa entender bem quais as entradas pedidas e como formatar os dados

Visualização de Dados

- Mas, se souber python, é possível até mesmo importar uma planilha direto do Google Drive e fazer seus plots com os tutoriais do link anterior

Validando os dados coletados: Testes Estatísticos

Testes Estatísticos

- Testes estatísticos permitem validar, com uma porcentagem de confiança aceitável, se dois grupos de dados são diferentes o bastante
 - ◆ Muito útil para provar que nova abordagem é de fato melhor que controle
 - ◆ Ou, até, que a nova abordagem não difere de um “ground truth”

Testes Estatísticos

- Para isso, as amostras das populações a serem testadas precisam ser representativas o bastante
 - ◆ Idealmente, pelo menos 30 de cada grupo (número mágico relativamente aceito. Porém, pode ser encontrado com teste de poder estatístico)
 - ◆ Vai depender do contexto da aplicação

Testes Estatísticos

→ Um dos jeitos de achar o valor adequado:

◆ Teste de poder estatístico

- d de Cohen (tamanho do efeito)
- Poder estatístico
- alfa (p-valor) desejado
- Resultado é o tamanho mínimo da amostra

→ Em Python:

```
result = (...).solve_power(effect_size=0.8, power=0.8, alpha=0.05)
```


Testes Estatísticos

- Com uma amostra boa o bastante, é preciso verificar a normalidade da amostra
 - ◆ Existem diversos testes estatísticos para amostras que seguem a distribuição normal
 - ◆ E outros, diferentes, para aquelas que não seguem
- Então, é preciso saber se dados são dependentes ou não
- E, por fim, quantas amostras deseja-se comparar

Testes Estatísticos

- Normalmente, independente do teste estatístico aplicado, buscamos um valor de alfa = 0.05 (5%)
- ◆ E, portanto, um p-valor < 0.05 para rejeitar a hipótese nula que as amostras são iguais
- ◆ Isso para uma cauda
 - Normalmente temos a intuição de qual amostra é maior, e verificamos se estatisticamente é, de fato, maior. Ou é igual.

Testes Estatísticos			
Paramétricos		Não-Paramétricos	
Independentes	Vinculados	Independentes	Vinculados
2 amostras	2 amostras	2 amostras	2 amostras
Teste <i>t</i> (Student)	Teste <i>t</i> (Student)	Mann-Whitney	Wilcoxon
		T. da Mediana	T. dos sinais
		χ^2 (2 x 2)	Mac Nemar
		Proporções	Binomial
		Exato (Fisher)	
Mais de duas	Mais de duas	Mais de duas	Mais de duas
Análise de variância	Análise de variância	Kruskal-Wallis	
		Mediana (m x n)	Cochran
		χ^2 (m x n)	Friedman
		Nemenyi	

Escolha de teste estatístico

Fonte: http://www.forp.usp.br/restauradora/gmc/gmc_livro/gmc_livro_cap14.html

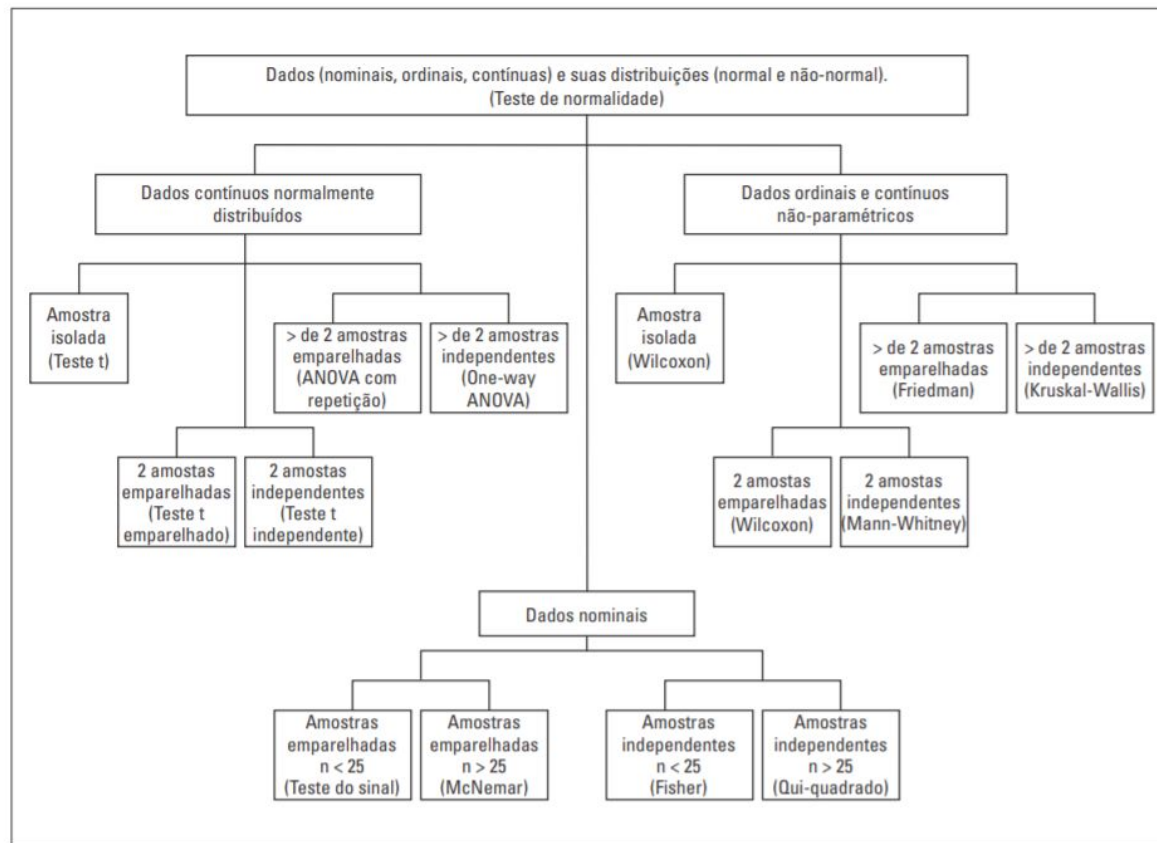
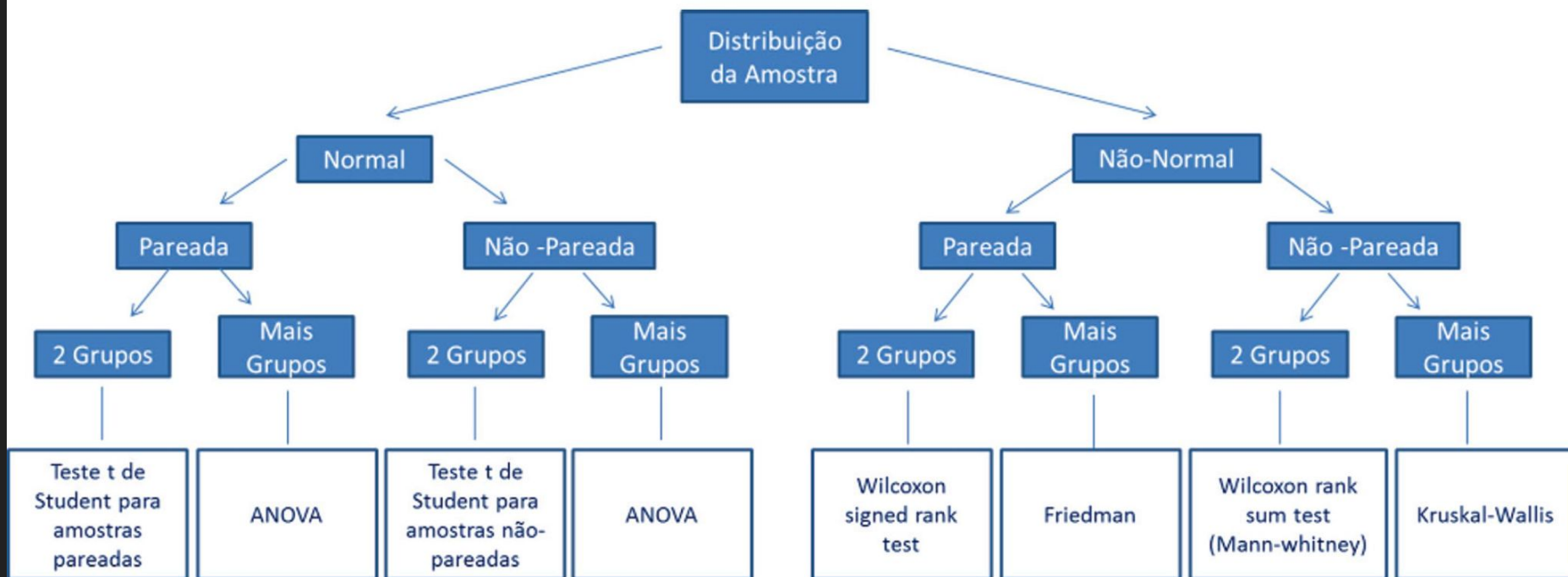


FIGURA 2 - Fluxograma esquematizando os passos na seleção de um determinado teste estatístico.

Escolha de teste estatístico

Fonte: An, Tien & Cuoghi, Osmar. (2004). A utilização da estatística na Ortodontia. Revista Dental Press De Ortodontia E Ortopedia Facial. 9. 10.1590/S1415-54192004000600014.

Esquema 1



Escolha de teste estatístico

Fonte: Lopes, Bernardo et al. Bioestatísticas: conceitos fundamentais e aplicações práticas. Revista Brasileira de Oftalmologia [online]. 2014, v. 73, n. 1 [Acessado 23 Outubro 2021], pp. 16-22.
Disponível em: <<https://doi.org/10.5935/0034-7280.20140004>>. ISSN 1982-8551.

Exemplos em Python

Referências

Referências

- [1] Jameson, A & Jacko, J & Sears, A. (2002). Adaptive Interfaces and Agents.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.91.4188&rep=rep1&type=pdf>
- [2] Joshi, A.; Kale, S.; Chandel, S.; and Pal, D. K. 2015. Likertscale: Explored and explained. British Journal of Applied Science & Technology 7(4): 396
<https://medium.com/@alisson.rlima/manipula%C3%A7%C3%A3o-de-planilhas-com-google-colab-google-sheets-e-pandas-6565a5a89826>

Referências

→ Visualização

- ◆ <https://towardsdatascience.com/introduction-to-data-visualization-in-python-89a54c97fbed>
- ◆ <https://www.data-to-viz.com/>

Referências

→ Estatísticas:

- ◆ https://en.wikipedia.org/wiki/Effect_size
- ◆ <https://machinelearningmastery.com/statistical-power-and-power-analysis-in-python/>
- ◆ <https://statsthinking21.github.io/statsthinking21-python/09-StatisticalPower.html>
- ◆ <https://www.statisticshowto.com/cohens-d/>
- ◆ <https://machinelearningmastery.com/a-gentle-introduction-to-normality-tests-in-python/>