

PROJETO 2

ANÁLISE INFERENCIAL – COMPARAÇÃO DE DUAS MÉDIAS

TESTE A/B -

Regras:

1. O Projeto 2 é de até 4 integrantes na equipe. São permitidas equipes com membros de salas diferentes. Se você fizer uma aula *studio* em uma turma que não a sua avise ao seu professor para não ficar com falta.
2. Elaborar um relatório que contenha as análises descritiva e inferencial que contribuíram para avaliar se as interfaces A e B se diferenciam com relação ao tempo descrito até cumprimento de determinada tarefa.
3. O relatório deve ser feito em IPython Notebook (com células de texto formatadas em Markdown quando conveniente) e deve conter, pelo menos, as seguintes seções:
 - a. Introdução (descrição das interfaces A e B (o que diferenciam) e tarefa em comum nas duas interfaces),
 - b. Análises descritiva e inferencial e explicação dos resultados,
 - c. Conclusão.

A estrutura do documento deve ser clara e de fácil compreensão da linha de raciocínio.
4. Deve ser colocado no github de cada integrante do grupo
5. O site de teste deve ser deixado no ar no Firebase Hosting até o fim do semestre
6. O arquivo DEVE ser com extensão .ipynb

Cronograma:

Turmas A, B e C

DATA	EM SALA – Aula Studio
03/11	<p>Páginas das interfaces A e B devem já estar no ar sendo acessadas para mensuração do tempo (ou outra métrica) até cumprimento da tarefa traçada pela equipe. O professor precisa conhecer o link até 23:59 de quinta-feira</p> <p>O grupo precisa ter claro:</p> <ul style="list-style-type: none">• Qual a tarefa realizada pelo usuário?• Como ela será comparada nas versões A e B?
07/11	<p>Entrega da técnica inferencial que será utilizada pela equipe. Essa entrega deve conter uma justificativa da escolha da técnica inferencial e ainda um exemplo com todas as contas feitas e uso do resultado para tomada de decisão.</p>

	<p>Esse exemplo pode ser tempos dos acessos das páginas considerando uma subamostra ou com valores totalmente hipotéticos ($n = 6$ de cada interface por exemplo é suficiente nesse exemplo).</p> <p>Aqui, o objetivo é que vocês já saibam a técnica inferencial que irão utilizar mesmo enquanto a página está no ar coletando dados no firebase.</p>
17/11	<p>Entrega Final do Projeto 2</p> <p>Essa entrega deve conter uma introdução detalhada sobre a tarefa e a diferença entre as interfaces, nomeadas pelo teste de A e B.</p> <p>Deixe claro que tipo de teste seu grupo fez e as hipóteses testadas.</p> <p>Exemplos:</p> <ul style="list-style-type: none">• Estamos comparando diferença de médias?• As distribuições das populações que geraram as amostras A e B são conhecidas? São iguais?• A variância de cada uma das populações é igual?• Qual o teste usado? Qual a estatística de teste? <p>Análise descritiva e inferencial cuidadosa com os dados que refletem todos os acessos enquanto as páginas ficaram no ar.</p> <p>Conclusão final.</p>

Observações:

Bootstrap e teste de permutação

Testes não paramétricos e baseados em dados (sem suposições sobre a população) só podem ser usados se o grupo já fez um teste paramétrico, para valor comparativo Exemplos de testes não paramétricos: bootstrap e teste de permutação.

Os testes paramétricos são aqueles baseiam em distribuições com formulas matemáticas conhecidas, por exemplo: teste z, teste t, teste t de Welch, etc.

Tarefa da interface

Procure definir a tarefa do usuário como algo que realmente permita avaliar se a web app funciona no geral. Por exemplo, se a tarefa for “ache o projeto 1 de ModSim” pode ser que este projeto em particular esteja bem posicionado.

Por outro lado, se você alternar entre alguns objetivos fornecidos ao usuário, poderá ter uma medida mais representativa.

Uma maneira fácil de fazer isso é usar a página de sorteio e encaminhar o visitante para uma dentre várias versões (quase iguais) do site.

Bibliografia sugerida:

Montgomery, D. **Engineering Statistics**, 5th Ed.

Chap.5 – Decision making for two samples

Magalhães e Lima. **Noções de Probabilidade e Estatística**.

Cap. 9. Especialmente seção 9.2 e tabela 9.1

RUBRICS DE AVALIAÇÃO DO OBJETIVO DE APRENDIZADO

Objetivo de aprendizado	Insatisfatório (I)	Em desenvolvimento (D)	Essencial (C)	Proficiente (B)	Avançado (A)
Planejar um teste A/B que permita avaliar quão bem um app cumpre seu objetivo	Apresentou entregas insuficientes ou atrasadas.	Conseguiu colocar o site no ar, mas falta coleta de dados ou faltam mudanças que permitem caracterizar as versões A e B	<p>As versões A e B do site existem.</p> <p>Os dados coletados têm as informações necessárias para o teste estatístico.</p> <p>Apresentaram um notebook que trata os dados e faz análise descritiva</p> <p>A tarefa a executar é deixada clara para o usuário</p>	Montou o teste de forma que realmente meça o desempenho do website de forma significativa, eventualmente variando detalhes das tarefas específicas a serem realizadas.	<p>Realizou os comportamentos de B e C de maneira excelente</p> <p>Procurou variar métricas ou incorporar tarefas mais sofisticadas</p>
Execução do teste A/B.	Não há nada que dê indicativos de construção de uma base de dados.	Não fica claro qual a métrica escolhida para avaliar possíveis evidência de diferença entre as interfaces.	<p>Usa função pronta do teste de hipóteses já fornecida no Python.</p> <p>Técnica inferencial não está adequada e/ou está incompleta (não descrevi as hipóteses a ser testada, falta estatística de teste, e/ou falta suposições necessárias)</p> <p>Não deixa claras as suposições feitas para escolha da técnica.</p>	<p>Técnica inferencial está adequada e bem escrita (descrevi as hipóteses a ser testada, estatística de teste, suposições necessárias).</p> <p>Escreve o código referente ao teste inferencial escolhido para a análise (as funções do scipy.stats que não resolvem o teste sozinhas podem ser usadas).</p> <p>Análise descritiva e/ou inferencial estão incompletas/erradas.</p> <p>Conclusão em termos do problema está incompleta/errada.</p>	<p>Realizou os comportamentos de B e C de maneira exemplar.</p> <p>Realizou comparação com outras técnicas de análise: intervalos de confiança, testes bootstrap e de permutação</p>