Universidade Federal do Rio de Janeiro

Departamento de Ciência da Computação

Trabalhando com Grande Volume de Dados

MAB102

**Aplicações de Estatística**

**Nome:** Patrícia de Andrade Kovaleski

**DRE:** 113090316

**Questão 1:** Nova feature para um site

Descrição:

Um portal de notícias implementou um play automático de vídeos em suas páginas, visando manter seus usuários por mais tempo no site. Deseja-se descobrir se vale a pena manter esta feature pois seu custo por visualização é de 0.005 centavos, e cada minuto adicional de navegação gera em média 5 centavos de lucro.

Dados:

* *Populacao\_tempo.csv* - dados contendo os tempos de acesso de um mês anterior a implementação da feature.
* *Amostra\_tempo.csv* - dados contendo os tempos de acesso do último mês, após a implementação feature.

Hipóteses:

Hipótese Nula:

H0: μ = μ0

Hipótese Alternativa:

H0: μ ≠ μ0

Procedimento:

Para determinarmos se vale a pena ou não manter a nova feature no site precisamos primeiro verificar se houve alguma mudança significativa na duração média de cada acesso. Dessa forma, podemos calcular o impacto econômico dessa mudança baseado nos custos de sua manutenção e o lucro gerado pelos minutos a mais de visualização.

Porém, mesmo que haja uma diferença entre as médias de acesso, seja ela positiva ou negativa, ainda é preciso verificar se de fato estamos avaliando uma população diferente da original. Ou seja, é preciso avaliar se os dados coletados sobre os tempos de acesso após a implementação da feature são diferentes por pertencerem a uma nova realidade do problema ou se apenas representam uma das variações da situação original.

Respostas:

Após analisar todos os dados contidos nos respectivos *datasets*, as médias calculadas foram:

Média da população = 1.99830

Média da amostra = 3.47206

De acordos com os valores acima, vemos que houve um aumento na duração média de cada acesso após a implementação da feature.

A nova feature deve ser mantida pois gera 2.36880648525 centavos de lucro!

**Questão 2:** Clicks do site

* Uma métrica comum em sites de e-commerce é o número de clicks que um usuário efetua durante a navegação. Um grupo de marketing quer fazer uma campanha de um novo produto, entretanto não sabe se apresenta ele apenas na página ou como um pop-up, que toma a conta de toda a tela. Visando responder esse problema, dois grupos foram selecionados. Para o primeiro grupo, foi apresentado apenas a tela com o produto. Para o segundo, foi apresentado a tela com o pop-up.
* Utilize técnicas estatísticas para informar se faz diferença utilizar o pop-up ou não.

**Questão 3:** Produtor de cinema

* Um produtor acredita que deve-se construir um poster para um filme utilizando a maior quantidade de atores possíveis. O fato se deve a uma correlação existente entre a nota no IMDB (imdb\_score) e o número de faces existentes nos posters (facenumber\_in\_poster).
* Verifique a existência dessa correlação e tente ajudar o produtor a conseguir o melhor imdb\_score possível.
  + Arquivo:
    - movie\_metadata.csv