

Trabalho Teste de Hipótese

Joao Vitor Elias Carvalho - 112037860

October 3, 2016

1 Primeiro Problema

1.1 Descrição do Problema

Um portal de notícias implementou um play automático de vídeos em suas páginas, visando manter seus usuários por mais tempo no site. Analise os dados históricos contendo os acessos do último mês, e mostre se houve um aumento significativo de tempo de navegação após a implementação da feature.

Sabendo que o custo por visualização desse play é de 0.005 centavos, e que cada minuto adicional de navegação gera em média 5 centavos de lucro, defina se essa nova feature deve ser mantida ou não.

1.2 Método

Esse é o problema em saber se uma amostra veio de uma certa população. Para isso vamos utilizar um método simples de teste de hipótese. Considerando que minha hipótese nula é: $\mu \geq \mu_0$ e minha hipótese alternativa é: $\mu < \mu_0$, podemos saber isso descobrindo aonde na distribuição normal aquela amostra se encaixa. Para isso vamos utilizar a fórmula:

$$z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$$

Sabendo isso podemos descobrir qual nível de significância possui atribuído aquele ponto na curva. Ou seja, qual a área a partir daquele ponto até o infinito. Ou o complemento da área de menos infinito até aquele ponto.

1.3 Resultado

Após computar o desvio padrão da população, média populacional e média amostral obtive os resultados de: $z = 23.3587947799$ com um valor associado de p-value de: $pvalue = 0.0$. Com isso podemos concluir com altíssimo grau de certeza que a amostra não veio da mesma população e a feature deve ser mantida.

2 Segundo Problema

2.1 Descrição do Problema

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.

2.2 Método

Esse é o problema em saber se duas amostras vieram da mesma população. Pra isso possuímos dois tipos de testes diferentes: Qui-Quadrado para duas amostras e Kolmogorov-Smirnov para Duas Amostras. Nesse caso vou utilizar o teste de Kolmogorov-Smirnov como é um teste bem simples e consegue bons resultados.

O teste de Kolmogorov-Smirnov para Duas Amostras utiliza a função de distribuição acumulada das duas amostras e descobre o ponto em que a diferença entre elas é maior. Então utiliza isso como estatística de teste.

Vamos utilizar 0 para quando não houve click e 1 para quando houve click para pré-processar nossos dados.

2.3 Resultado

Após computar o teste para as duas amostras obtive os seguintes valores: $Ks = 0.077723078041$ e $pvalue = 0.00457508275548$. Além disso obtive a soma de clicks de 301 para a amostra A e 387 para a amostra B. Com esse valor de pvalue pequeno podemos concluir que as duas amostras não vieram da mesma população. A amostra B possuiu mais cliques que a amostra A, então sugiro usar a pop-up para apresentar a campanha.

3 Terceiro Problema

3.1 Descrição do Problema

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 e o número de faces existentes nos posters . Verifique a existência dessa correlação e tente ajudar o produtor a conseguir a melhor nota possível.

3.2 Método

Podemos descobrir se existe correlação entre o número de faces existente nos posters e a nota utilizando um teste como o de Spearman. Esse teste consegue identificar correlações entre dados e não apenas lineares com o de Pearson.

3.3 Resultado

Após computar o teste obtive os seguintes resultados: $\rho = -0.0868173370519$ e $pvalue = 9.07119693373e - 10$. Com isso podemos concluir que existe uma relação inversamente proporcional entre os dados com um grau de certeza alto.