**Relatório da Reprodução da Análise Quantitativa de um Experimento**

**Elissandra G. Pereira, Victor Luis Perszel**

Departamento de Informática – Universidade Federal do Paraná (UFPR)  
Rua Evaristo F. F. da Costa, 418 - Jardim das Américas  
80050-540 - Curitiba (PR) - Brasil

{egpereira,vlperszel}@inf.ufpr.br

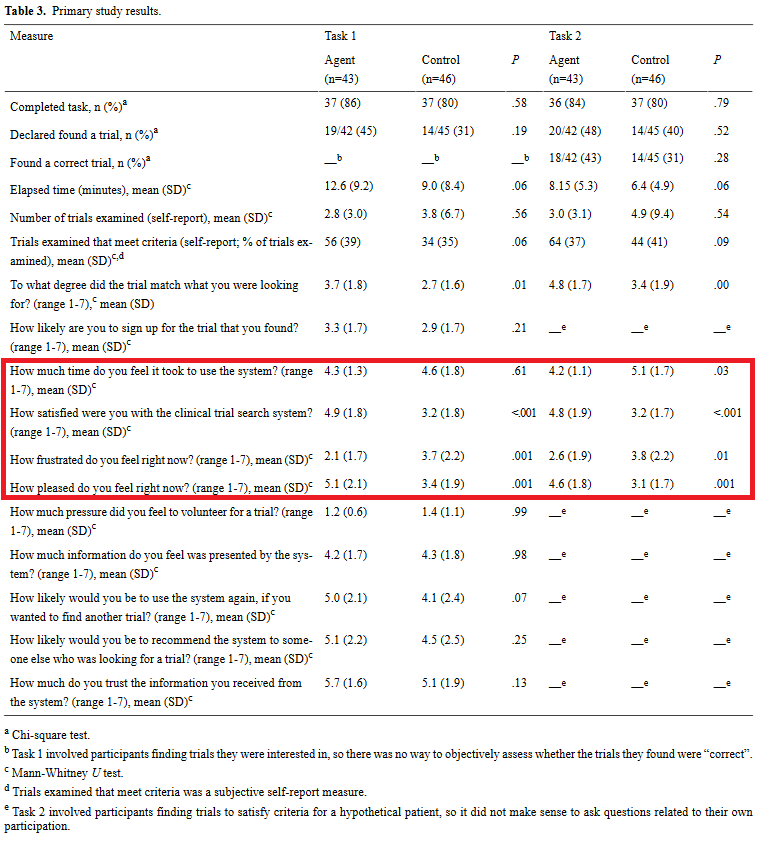
***Abstract.*** *This paper reports the reproduction of the quantitative analysis of the experiment identified in the paper “Improving access to online health information with conversation agents: a randomized controlled experiment”, describing each step performed during the tests in the statistical software and showing the result achieved*

***Resumo.*** *Este artigo relata a reprodução da análise quantitativa do experimento identificado no artigo “Improving Access to Online Health Information With Conversational Agents: A Randomized Controlled Experiment”, descrevendo cada etapa executada durante os testes no software estatístico e mostrando o resultado alcançado.*

1. **Identificação do Artigo**Para realizar o trabalho procuramos um artigo avaliando de forma comparativa a usabilidade de dois ou mais elementos, através de um teste com o usuário. O artigo precisava incluir os dados brutos do experimento e seu resultado qualitativo, para que pudéssemos tentar reproduzí-lo.

O artigo escolhido foi “*Improving Access to Online Health Information With Conversational Agents: A Randomized Controlled Experiment*” (Bickmore, Utami, Matsuyama e Paasche-Orlow 2016), que faz uma análise de usabilidade entre um novo agente para pesquisa de informações de saúde (*agente)* e a ferramenta convencional *(controle)*, através de um experimento controlado com usuário.

1. **Reconhecimento do Testes Estatísticos do Experimento**O artigo escolhido trabalha com a análise de diferentes critérios de avaliação da ferramenta testada, utilizando o chi-quadrado e o Mann-Whitney para criação de resultado estatístico. Gerando os resultados na tabela [1].  
    Para a reprodução escolhemos os tópicos em destaque na imagem:*“How much time do you feel it took to use the system?*” (Quanto tempo você acreditar ter levado para usar o sistema?), *“How satisfied were you with the clinical trial search system?”* (Quão satisfeito você ficou com o sistema de pesquisa clínico?), *“How frustrated do you feel right now?”* (Quão frustrado você se encontra nesse momento?) e *“How pleased do you feel right now?”* (Quão satisfeito você se encontra nesse momento?).  
    Essa escolha se deve para não nos estendermos neste artigo e pois são os dados que refletem sobre a questão de usabilidade, que foi nosso objetivo com a pesquisa feita anteriormente.

 **Tabela 1**: resultados obtidos pelo autor em seu artigo. Em destaque os tópicos escolhidos para nossa reprodução.

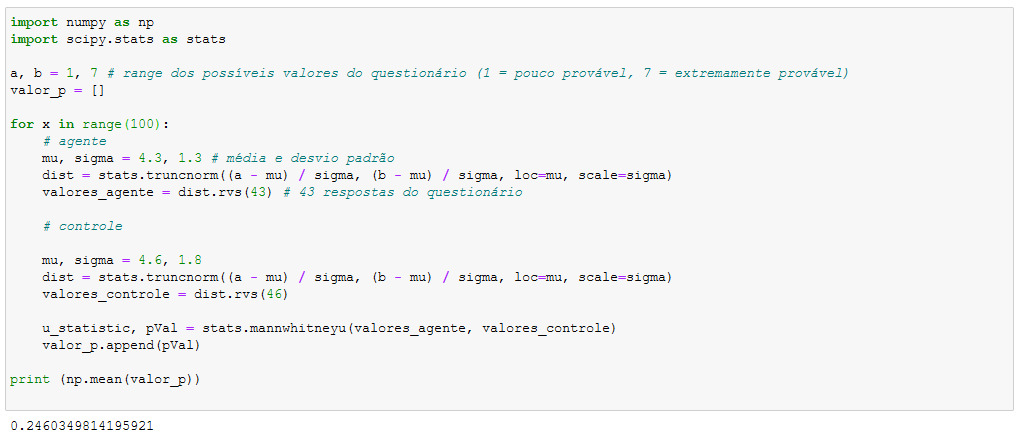
1. **Reprodução utilizando o software estatístico**

Para simular os resultados qualitativos do experimento analisado foi empregada a linguagem Python.   
 Como o artigo não fornece todas respostas do questionário (seria até inviável para o pesquisador colocar todos esses números no texto) geramos valores aleatórios com base na média, desvio padrão e número de respostas obtidas, tanto para o agente quanto para o controle. Com isso, calculamos o valor-p para que possamos comparar com o resultado do autor do artigo e tirar nossas conclusões. Já que, por azar, valores aleatórios podem causar discrepâncias muito grandes, rodamos o algoritmo descrito para a mesma questão cem vezes e tiramos a média de todos os valor-p, para assim podermos termos uma análise melhor para com o resultado descrito no artigo.

1. **Resultado da Reprodução**

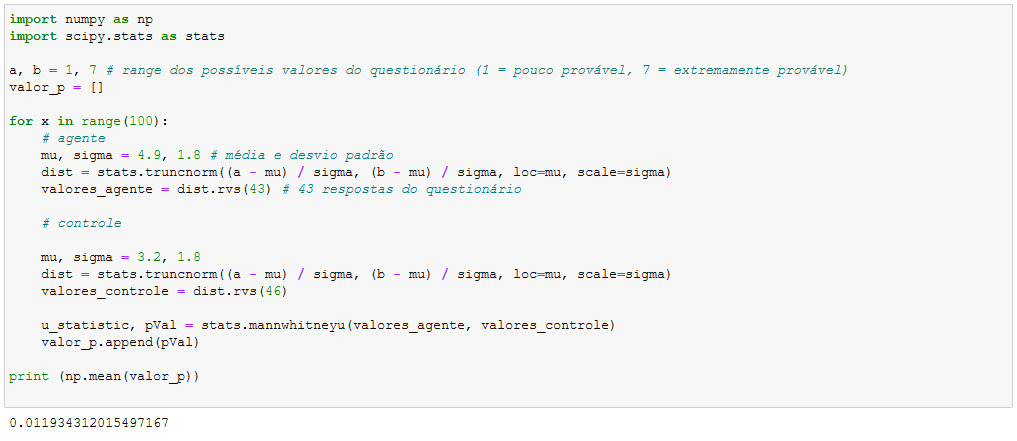
Através dos valores gerados, chegamos ao seguinte resultado como reprodução:

Aqui analisamos os resultados primários da questão: *"How much time do you feel it took to use the system?"*



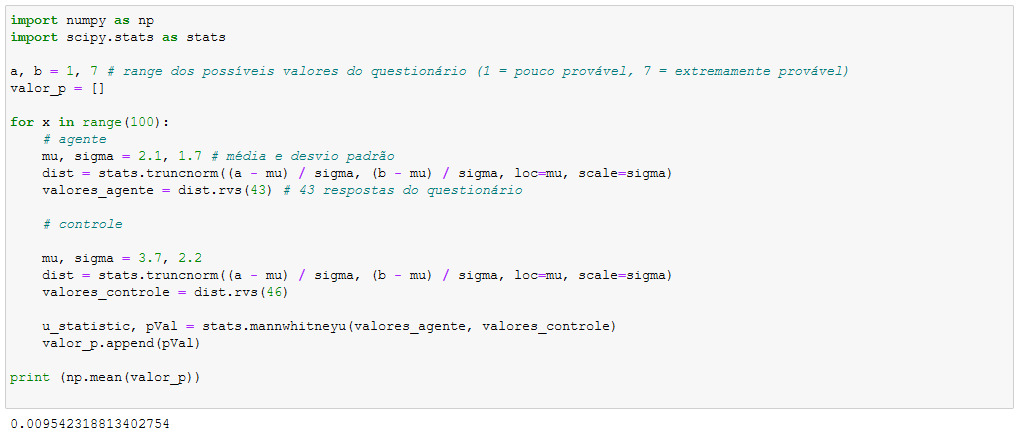
**Imagem 1**: resultados obtidos durante a reprodução da análise de "How much time do you feel it took to use the system?”

Os valores para o agente foram de média 4,3 e desvio padrão 1,3 (com 43 dados). Já para o controle é indicado média 4,6 e dp 1,8. Com esses dados disponíveis, o artigo indica ter chego em um valor p de 0,61 (com 46 dados). O nosso resultado obtido para o valor-p foi de 0,24. Apesar da diferença, ambos valores indicam um número acima de 0,05, ou seja, aceitando a hipótese nula (H0), indicando não haver diferença entre os resultados.

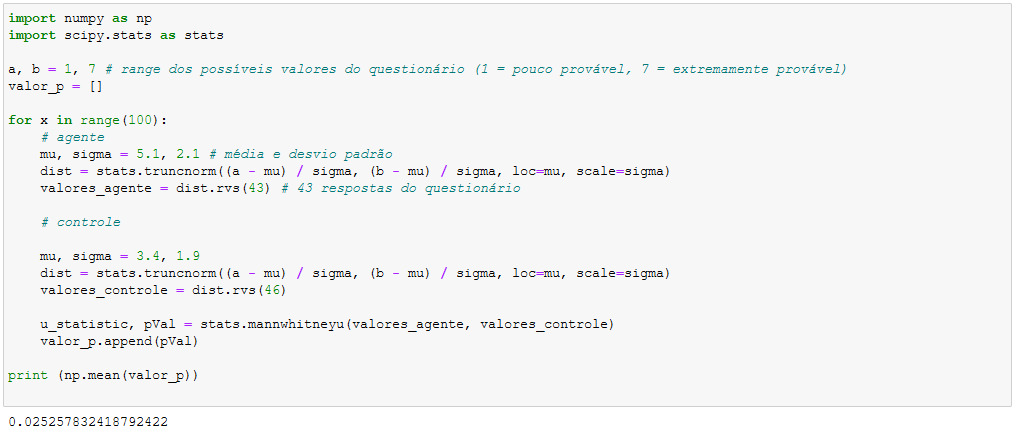
Na questão *"How satisfied were you with the clinical trial search system?"* **Imagem 2**: resultados obtidos durante a reprodução da análise de *"How satisfied were you with the clinical trial search system?*

Temos para o agente a média de 4,9 e dp 1,8, enquanto para o controle é mostrado média 3,2 e dp 1,8. Com isso, é apresentado um valor-p <001. Nossos resultados indicam o valor exato 0,01, em concordância com o mostrado pelo artigo e rejeitando a hipótese H0.

Sobre o tópico *“How frustrated do you feel right now?”*

**Imagem 3**: resultados obtidos durante a reprodução da análise de *“How frustrated do you feel right now?”*

O artigo original apresenta agente com média 2,1 e desvio 1,7 e controle com esses valores 3,7 e 2,2, respectivamente, mostrando um valor-p de 0,01. Nossa reprodução chegou no valor-p de 0,009, indicando uma aproximação satisfatória com o resultado original.

Sobre a questão *“How pleased do you feel right now?”*

Os dados originais apresentam uma média de 5,1 e desvio padrão de 2,1 para o agente, enquanto o controle apresentou os valores de 3,4 e 1,9 para os mesmos cálculos. O valor-p descrito com isso foi de 0,01. Nossa reprodução do valor-p foi 0,025. Apesar da discrepância, ambos apresentam valor < 0.05, ou seja, a hipótese nula foi rejeitada em ambos casos.

Apesar das limitações criadas pela ausência dos dados brutos, o resultado da reprodução conseguiu replicar de forma satisfatória os resultados, mesmo quando houve uma certa diferença entre os valores obtidos. O exercício permitiu criarmos de forma satisfatória os valores quantitativos de resultado de um experimento controlado.

1. **Referência**Timothy W Bickmore, Dina Utami, Robin Matsuyama, Michael K Paasche-Orlow. (2016) “Improving Access to Online Health Information With Conversational Agents: A Randomized Controlled Experiment’. Originally published in the Journal of Medical Internet Research .