

L3P1: Experimento com Jukeboxes

Disciplina: Fundamentos de Pesquisa em Ciência da Computação 2 (FPCC2)

Professor: Nazareno Andrade

Aluno: Júlio Barreto Guedes da Costa

Os dados

Os dados utilizados nesta análise foram coletados experimentalmente, conforme reportado [neste artigo](#), em que esta análise se baseia. Os dados reportam a satisfação de um indivíduo enquanto usuário de um sistema de compartilhamento de música no ambiente, atestando a diferença na satisfação quando o mecanismo de seleção das músicas é alterado. Para um conjunto de 23 indivíduos, foi definido um mecanismo padrão, nomeado *baseline*, e três propostas de novos mecanismos:

- *Like/Dislike*: Nesse mecanismo, cada usuário pode marcar que gostou ou não gostou de uma música, mas o mecanismo não controla a reprodução da música atual ou das próximas, agindo apenas como uma forma de *feedback* para o usuário que adicionou a música.
- *Up/downvoting*: Nesse mecanismo, cada usuário pode marcar uma música na lista de músicas a serem reproduzidas a seguir com um voto positivo ou negativo. A medida que os usuários votam nas músicas, a música melhor votada no momento é escolhida como a próxima a ser reproduzida.
- *Skip*: Nesse mecanismo, cada usuário pode marcar se quer ou não pular a música que está sendo reproduzida no momento. Caso o número de votos para pular a música atinja um certo limite, a música é pulada, passando para a próxima música da lista.

Além desses três mecanismos, também foi proposto o mecanismo *Combined*, que une as funcionalidades dos três mecanismos simultaneamente.

Analisando os dados

Analisando os dados do artigo, queremos entender:

1. Qual o tamanho do efeito de cada uma das alternativas do sistema comparada com o *baseline*? Esse efeito é relevante? Vale a pena para um desenvolvedor implementar uma delas?
2. Qual o tamanho do efeito da combinação de alternativas (a *combined*) em relação às alternativas separadas?
3. Considerando a escolha entre *skip* e *up/downvoting*, você recomendaria a alguém interessado em aumentar a satisfação de usuários que se prefira a implementação de um deles?
4. Se ela é mais trabalhosa, em que situação você recomendaria a implementação das várias alternativas juntas em um sistema?

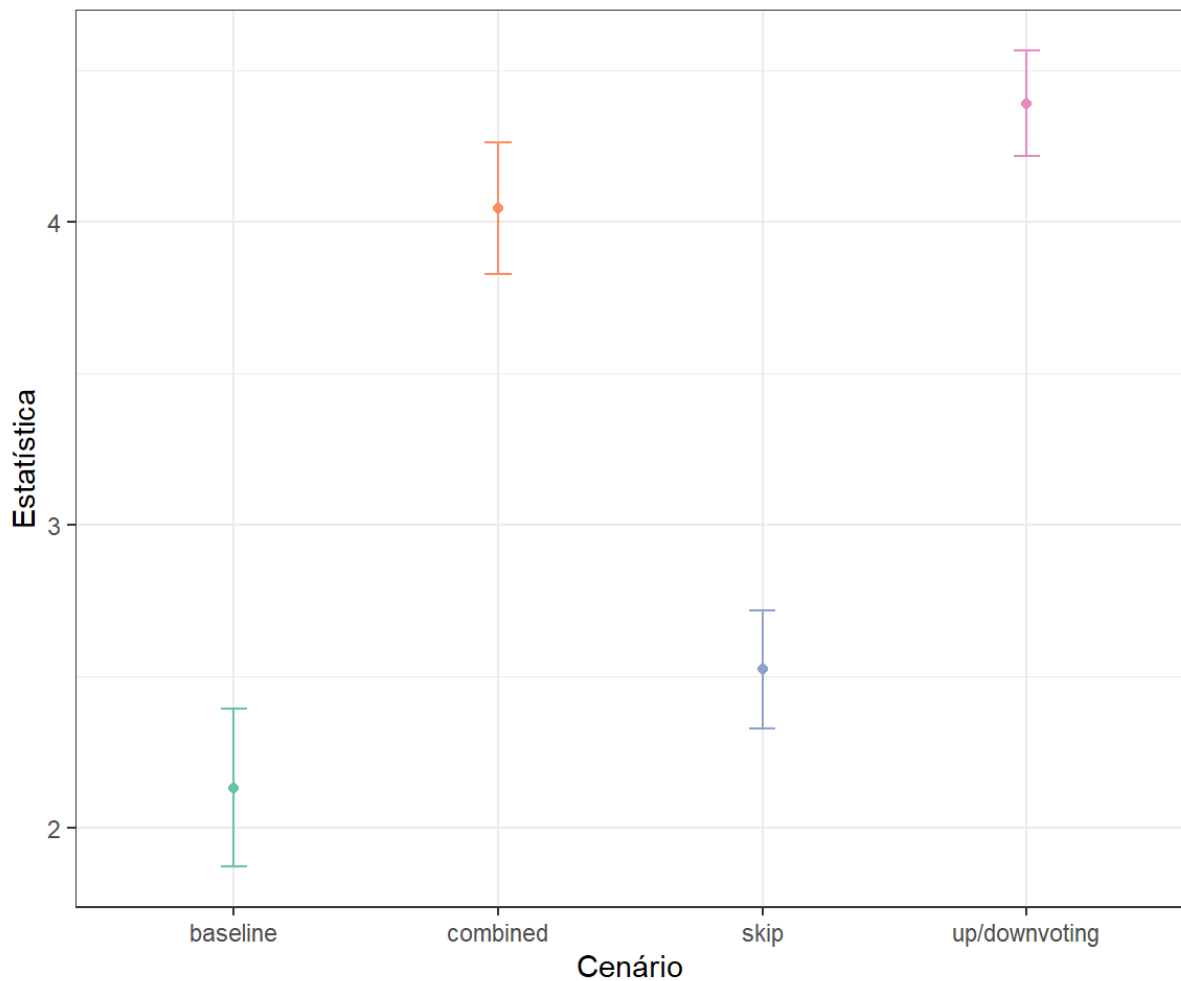
Por questão de simplificação, omitiremos o mecanismo *Like/Dislike* nesta análise.

A scatter plot showing satisfaction levels for four scenarios: baseline, combined, skip, and up/downvoting. The y-axis represents satisfaction from 1 to 5. Each scenario has a cluster of small dots and one larger mean dot.

Cenário	Satisfação (Mean)	Satisfação (Range)
baseline	2.1	1.0 - 3.0
combined	4.0	2.5 - 5.0
skip	2.5	1.5 - 3.0
up/downvoting	4.4	4.0 - 5.0

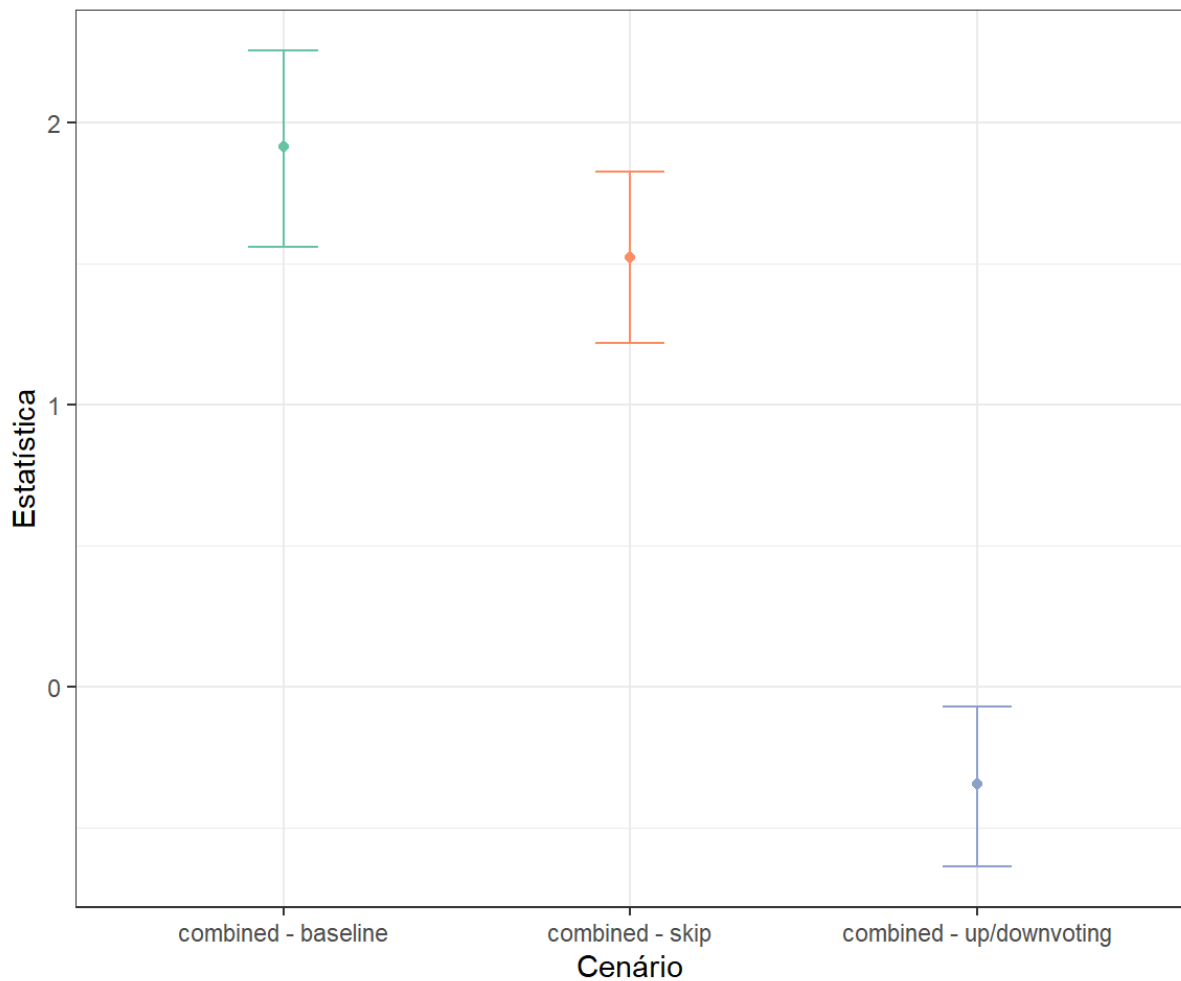
- *Baseline*: 2.13
- *Up/Downvoting*: 4.39
- *Skip*: 2.52
- *Combined*: 4.04

Entretanto, devido ao número limitado de indivíduos no sistema, é difícil entender com precisão o efeito da alteração do mecanismo na satisfação dos usuários. Assim, usando a técnica *bootstrapping*, realizaremos amostragens consecutivas de cada um dos cenários para entender melhor as estatísticas de cada cenário, estimando também os intervalos de confiança. O número de re-amostragens definido nesta análise foi 4000. Temos então a seguinte visualização:



Podemos ver que o *baseline* obteve os menores valores. Entretanto, o mecanismo *Skip* ainda possui uma intersecção com o *baseline*, não sendo possível dizer com certeza se existe melhoria de satisfação gerada com seu uso. Os outros mecanismos, *combined*, e *up/downvoting* mostraram melhoria relevante em relação ao *baseline*, e vale a pena que um desenvolvedor as implemente.

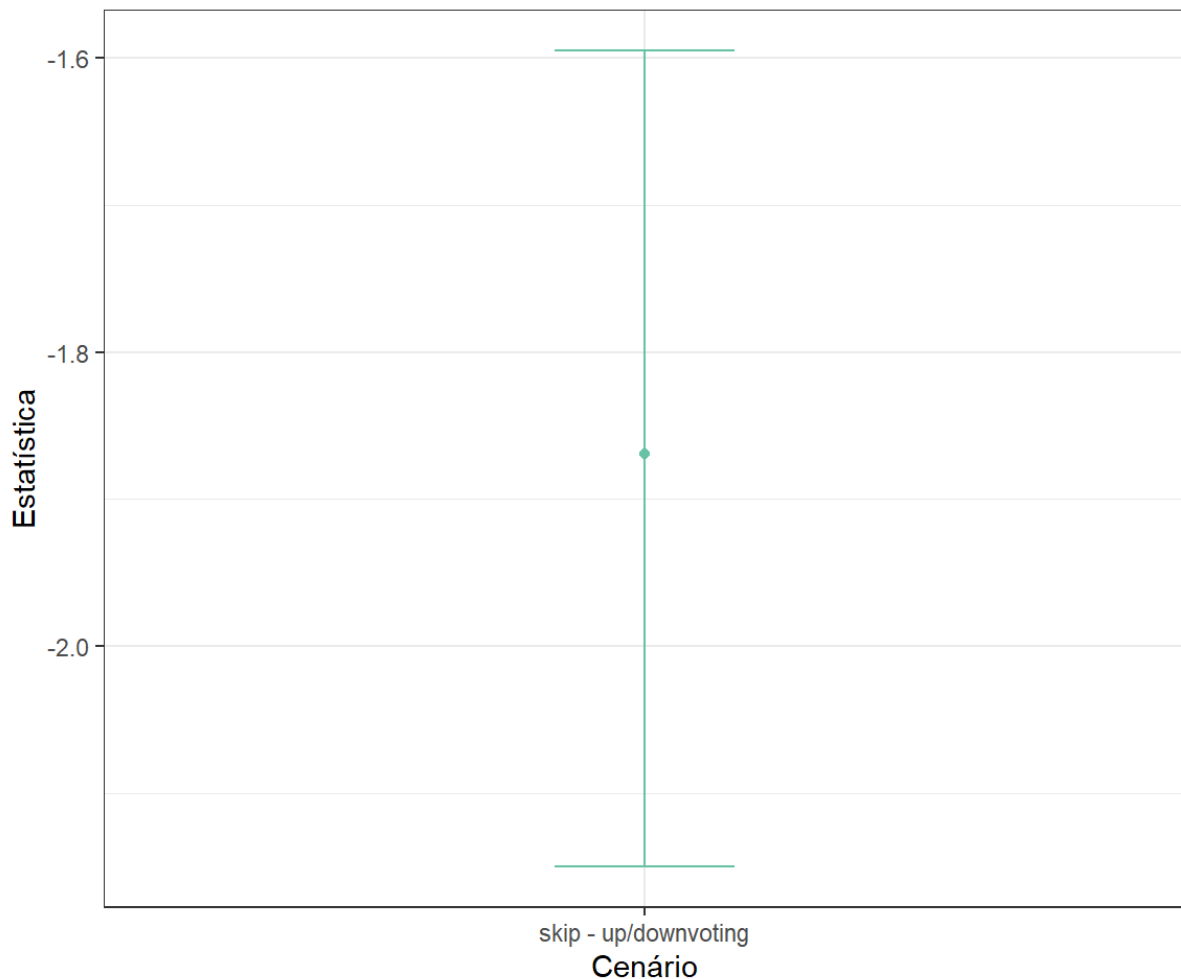
Para entender melhor a combinação dos mecanismos, *combined*, em relação à implementação de um único mecanismo, podemos calcular a diferença entre a satisfação dada à estratégia *combined* em relação à satisfação dos outros mecanismos. Assim, temos:



Pelo gráfico, podemos ver que a implementação de múltiplos mecanismos gera melhoria relevante, quando comparado com o mecanismo *baseline* e o mecanismo *skip*. Entretanto, quando comparado com o mecanismo *up/downvoting*, causa uma piora, porém não podemos dizer se essa piora é grande ou pequena. Em valores numéricos, temos:

- *combined - baseline*: média 1.91, IC [1.56, 2.25], com 95% de confiança;
- *combined - skip*: média 1.52, IC [1.22, 1.82], com 95% de confiança;
- *combined - up/downvoting*: média -0.34, IC [-0.64, -0.07], com 95% de confiança.

Quando tentando entender o mecanismo *skip* em comparação com o *up/downvoting*, podemos também calcular a diferença entre satisfação nos dois mecanismos:



Pela imagem, entendemos que o mecanismo *skip* gera uma satisfação inferior à gerada pelo *up/downvoting*, uma piora relevante. Em valores numéricos, temos que a média da diferença entre os dois mecanismos é -1.87, e IC [-2.15, -1.6] com 95% de confiança. Assim, considerando o pior desempenho do *skip*, o *up/downvoting* é preferível quando escolhendo entre implementar um deles.

Considerando a análise até então, podemos concluir que os mecanismos *up/downvoting* e *combined* apresentam melhoria relevante em comparação com os dois outros mecanismos. Porém, como o mecanismo *combined* possui um custo maior de implementação - já que é necessário implementar todos os outros mecanismos simultaneamente -, sua implementação só é preferível em substituição aos mecanismos *baseline* e *skip*; entretanto, para implementar o *combined*, é necessário implementar também o *up/downvoting*, que mostrou gerar uma satisfação superior, a implementação do *combined* só é interessante quando não for possível manter o *up/downvoting*.