

**UNIVERSIDADE FEDERAL DE ALFENAS - UNIFAL/MG**

**FERNANDA LARA VALADARES**

**MODELAGEM DE BATALHAS NO INSTAGRAM VIA DISTRIBUIÇÃO  
BETA E INFERÊNCIA BOOTSTRAP**

**ALFENAS/MG**

**2024**

**FERNANDA LARA VALADARES**

**MODELAGEM DE BATALHAS NO INSTAGRAM VIA DISTRIBUIÇÃO  
BETA E INFERÊNCIA BOOTSTRAP**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, área de concentração em Estatística Aplicada e Biometria da Universidade Federal de Alfenas-MG, como parte dos requisitos para a obtenção do título de Mestre. Linha de Pesquisa: Modelagem Estatística e Estatística Computacional.

Orientador: Prof. Dr. Eric Batista Ferreira

Coorientador: Prof. Dr. Carlos Pereira da Silva

**ALFENAS/MG**

**2024**

Sistema de Bibliotecas da Universidade Federal de Alfenas  
Biblioteca Central

Valadares, Fernanda Lara .

Modelagem de batalhas no Instagram via distribuição Beta e Inferência  
Bootstrap / Fernanda Lara Valadares. - Alfenas, MG, 2024.

62 f. : il. -

Orientador(a): Eric Batista Ferreira.

Dissertação (Mestrado em Estatística Aplicada e Biometria) -  
Universidade Federal de Alfenas, Alfenas, MG, 2024.

Bibliografia.

1. Campeonato. 2. Inferência. 3. Intervalo de confiança. 4. Queijos. 5.  
Sertanejos. I. Ferreira, Eric Batista, orient. II. Título.

FERNANDA LARA VALADARES

**MODELAGEM DE BATALHAS NO INSTAGRAM VIA DISTRIBUIÇÃO BETA E INFERÊNCIA  
BOOTSTRAP**

O Presidente da banca examinadora abaixo assina a aprovação da Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre em Estatística Aplicada e Biometria pela Universidade Federal de Alfenas. Área de concentração: Estatística Aplicada e Biometria

Aprovada em: 5 de abril de 2024.

Prof. Dr. Eric Batista Ferreira

Presidente da Banca Examinadora

Instituição: Universidade Federal de Alfenas - UNIFAL-MG

Profa. Dra. Adriana Dias

Instituição: Universidade Federal de Alfenas - UNIFAL-MG

Prof. Dr. Denismar Alves Nogueira

Instituição: Universidade Federal de Alfenas - UNIFAL-MG



Documento assinado eletronicamente por **Eric Batista Ferreira, Presidente**, em 05/04/2024, às 17:34, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [https://sei.unifal-mg.edu.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.unifal-mg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **1227399** e o código CRC **71459EB2**.

---

Dedico esta dissertação à minha amada filha, Ana Laura Valadares Ribeiro, e ao meu querido pai, Enes Cordeiro Valadares, que está sempre presente em meu coração, mesmo em sua ausência física. À Ana Laura, por ser minha fonte de inspiração e por seu constante apoio ao longo desta jornada acadêmica. E ao meu pai, cujo amor e sabedoria continuam a me guiar, mesmo que ele não esteja mais entre nós. Esta conquista é dedicada a vocês, com todo o meu amor e gratidão.

## AGRADECIMENTOS

Meu agradecimento ao meu estimado professor orientador, Dr. Eric Batista Ferreira, que se destacou não apenas como meu guia acadêmico, mas também como um mentor e amigo dedicado ao longo desta jornada de pesquisa. Sua orientação sábia, paciência incansável e apoio inabalável foram fundamentais para o sucesso desta dissertação. Expresso minha sincera gratidão por sua orientação, encorajamento e dedicação ao longo deste período desafiador. Seu apoio foi fundamental para alcançar este marco em minha jornada acadêmica.

À Universidade Federal de Alfenas (UNIFAL-MG), ao Programa de Pós-graduação em Estatística Aplicada e Biometria (PPGEAB), à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001

## RESUMO

Competições, concursos e campeonatos são comumente praticados nas mais diversas áreas - esportes, beleza, habilidades, música, etc - para comparar e ranquear competidores. Particularmente, alguns campeonatos seguem um esquema de chaves ou chaveamento, que podem ser simples, duplos ou múltiplos, para conduzir os vencedores até eleger o campeão. Nesses casos, o enfrentamento acontece aos pares, ou seja, quando dois oponentes se enfrentam, apenas um sai vitorioso. Quando não é possível haver empates ou repescagem, esses confrontos são chamados de “mata-mata”. Ao final de campeonatos desse tipo, o resultado é o agrupamento dos competidores nas classes: primeiro colocado, segundo, terceiros, quartos, quintos, etc. No entanto, como cada confronto acontece uma única vez, tem-se uma amostra de tamanho 1 de uma população conceitual de infinitas repetições daquele combate. Sendo assim, os agrupamentos resultantes de tais campeonatos possuem uma incerteza associada, que é normalmente desprezada. Seria útil modelar a incerteza envolvida em um campeonato, a fim de saber quais vencedores podem ser considerados realmente superiores e quais venceram por mero acaso. Para responder a essa demanda, este trabalho propõe uma abordagem metodológica que utiliza distribuição Beta, inferência Bootstrap paramétrico, e escala hedônica. Como exemplos de sua utilização, campeonatos foram conduzidos nos *stories* do Instagram, com os queijos populares no Brasil e os mais famosos cantores sertanejos. Desta forma, apresenta-se não somente uma forma de inferir sobre a igualdade dos competidores de cada batalha do campeonato, mas também a interpretação de quão bem aceitos eles são pelos seguidores que participaram das votações. Ao final, metodologia proposta se mostrou eficiente na modelagem da incerteza associada a agrupamentos promovidos por campeonatos em chaveamento, revelando que os queijos que se enfrentaram não podem ser ditos estatisticamente diferentes e que Marília Mendonça, Henrique e Juliano, Jorge e Mateus e Zé Neto e Cristiano são os cantores sertanejos preferidos pelos brasileiros.

Palavras-chave: Campeonato; Inferência; Intervalo de confiança; Queijos; Sertanejos.

## ABSTRACT

Competitions, contests and championships are commonly held in the most diverse areas - sports, beauty, skills, music, etc. - to compare and rank competitors. Particularly, some championships follow a key scheme, which can be single, double,..., to evolve the competitors until choosing the winner. In these cases, the confrontation takes place in pairs, that is, when two opponents face each other, only one emerges victorious. When it is not possible to have draws or repechage, these clashes are called “knockout”. At the end of championships of this type, the result is the grouping of opponents into classes: first place, second, third, fourth, fifth, etc. However, as each confrontation happens only once, we have a sample size 1 of a conceptual population of infinite repetitions of that combat. Therefore, the groupings resulting from such championships have an associated uncertainty, which is normally ignored. It would be useful to model the uncertainty involved in a championship, in order to know which winners can be considered truly superior and which won by mere chance. To answer this question, this work proposes a methodological approach that uses the Beta distribution and parametric Bootstrap inference. As examples, championships made on Instagram *stories* with popular cheeses in Brazil and country singers are used. In this way, the proposed methodology proved to be efficient in modeling the uncertainty associated with groups promoted by switching championships, revealing that the cheeses that faced each other cannot be said to be statistically different and that Marília Mendonça, Henrique and Juliano, Jorge and Mateus and Zé Neto and Cristiano are the country singers preferred by Brazilians.

Keywords: Championship; Inference; Confidence interval; Cheeses; “Sertanejos”.



## LISTA DE FIGURAS

1	Relacionamento da distribuição Beta com outras distribuições. . . . .	17
2	Formas simétricas (a) e assimétricas (b) da distribuição Beta. . . . .	18
3	Grafo em que os vértices representam pessoas e as arestas . . . . .	26
4	Escalas hedônicas de 9 pontos mista, nominal e de 3 pontos mista. . . . .	28
5	Restrição do espaço paramétrico. . . . .	36
6	Densidades Beta em confrontos com forças similares (a) e diferentes (b). . .	37
7	Resumo do procedimento metodológico proposto. . . . .	39
8	Sorteio dos confrontos para as chaves do campeonato dos . . . . .	41
9	Sorteio dos confrontos para as chaves do campeonato dos artistas sertanejos.	42
10	<i>Prints</i> dos <i>stories</i> anunciando o início das quartas de final (a) e a . . . . .	43
11	<i>Prints</i> dos <i>stories</i> utilizados nas quartas de final. . . . .	45
12	<i>Prints</i> dos <i>stories</i> utilizados nas rodadas semifinal e final. . . . .	46
13	Fluxo do campeonato dos queijos. . . . .	46
14	Intervalos de 95% de confiança de bootstrap para a força . . . . .	48
15	Grafo do resultado do campeonato dos queijos. . . . .	49
16	Grafo modificado do resultado do campeonato dos queijos. . . . .	50
17	Fluxo do campeonato dos sertanejos. . . . .	51
18	Exemplos de densidades Beta ajustadas a um confronto com forças restritas	53
19	Exemplos de comparações entre intervalos de 95% de confiança para as forças	54
20	Grafo com os artistas sertanejos, após o campeonato. . . . .	55
21	Grafo modificado do resultado do campeonato dos sertanejos. . . . .	55

## LISTA DE TABELAS

1	Escalas hedônicas de 3, 5, 7 e 9 pontos. . . . .	28
2	Foto utilizada, nome e fonte dos oito queijos utilizados neste trabalho. . . .	40
3	Dezesseis artistas sertanejos que figuravam na lista das cinquenta músicas mais . . . . .	42
4	Proporções observadas em todos os combates, suas respectivas forças restritas	47
5	Proporções observadas em todos os combates, suas respectivas forças restritas	52

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	12
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	13
2.1	CAMPEONATOS EM CHAVEAMENTO	13
2.2	A DISTRIBUIÇÃO BETA	16
2.2.1	Densidade, parâmetros e formas	17
2.2.2	Estimação dos parâmetros	19
2.3	INFERÊNCIA BOOTSTRAP	22
2.3.1	Bootstrap Não-Paramétrico	23
2.3.2	Bootstrap Paramétrico	24
2.4	GRAFOS	24
2.5	ESCALA HEDÔNICA	27
2.6	REDES SOCIAIS	30
2.7	PRODUÇÃO E CONSUMO DE QUEIJOS NO BRASIL	31
2.8	A MÚSICA SERTANEJA BRASILEIRA	33
<b>3</b>	<b>MATERIAL E MÉTODOS</b>	36
3.1	METODOLOGIA PROPOSTA	36
3.2	BATALHA DOS QUEIJOS	40
3.3	BATALHA DOS SERTANEJOS	41
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	45
4.1	BATALHA DOS QUEIJOS	45
4.2	BATALHA DOS SERTANEJOS	50
<b>5</b>	<b>CONSIDERAÇÕES FINAIS</b>	56
	REFERÊNCIAS	58

## 1 INTRODUÇÃO

Campeonatos organizados em *chaveamento* são aqueles que contêm confrontos diretos entre pares de oponentes, em uma ordem estabelecida por sorteio. Após cada confronto haverá um perdedor e um vitorioso. O vitorioso passa para a próxima fase, enquanto o perdedor sai do torneio. Seguindo esse esquema, a cada rodada, um conjunto de competidores sai do torneio. Os competidores que saem em uma dada rodada farão parte do mesmo grupo. Por exemplo, os dois competidores eliminados nas semi-finais, serão os terceiros colocados, fazendo parte, assim, de um mesmo grupo. Os dois que venceram, por sua vez, vão compor a final, e gerar o primeiro e o segundo colocados (grupos diferentes).

Mas, é possível afirmar que o primeiro colocado é melhor que o segundo? Em uma primeira análise, sim, por ele ter ganhado a final. Mas, ele não poderia ter vencido por acaso, por sorte? Sim, isso também pode acontecer. Então, como ter certeza que o primeiro é melhor que o segundo? Compreendendo o campeonato como uma amostra, não é possível ter certeza. Se um campeonato for entendido como uma realização de uma população conceitual de infinitos campeonatos semelhantes, ele pode ser dito uma amostra dessa população.

Sendo assim, os resultados variam de uma amostra para outra (de uma realização do campeonato para outra). Então, o que se pode estudar é a incerteza associada ao agrupamento gerado pelo campeonato, ou seja, a probabilidade de dois competidores diferentes estarem no mesmo grupo ou competidores estatisticamente iguais, serem classificados em grupos diferentes. Por exemplo, quando, ao final do campeonato, dois competidores ficam na terceira posição, qual é a probabilidade de eles serem estatisticamente diferentes (e este agrupamento estar errado)? Ou ainda, quando um competidor termina o campeonato em primeiro lugar e outro em segundo, qual é a probabilidade de eles serem, de fato, diferentes? Essa incerteza não se encontra bem posta na literatura.

Neste cenário, podem ser identificadas quantidades populacionais (parâmetros) e quantidades amostrais. Os parâmetros associados aos competidores, que indicam sua probabilidade de vitória, podem ser chamados de “força” do competidor. Os valores paramétricos dessa força podem ser estimados por meio de técnicas de Inferência Estatística, e esta pode ser potencializada por técnicas de computação intensiva, como o Bootstrap.

Assim, este trabalho tem por objetivo principal propor um método inferencial baseado na distribuição Beta e Bootstrap paramétrico para campeonatos do tipo “mata-mata”, em esquema de chaveamento duplo. Especificamente, são objetivos: i. Propor o conceito de força restrita de um competidor, com direta analogia à escala hedônica; ii. Estimar os parâmetros distribuição Beta de probabilidades via inferência Bootstrap paramétrico; iii. Ilustrar o método proposto com exemplos reais, uma batalha de queijos e uma batalha de cantores sertanejos, feitas nos *stories* do Instagram; iv. Propor opções gráficas de visualização dos resultados do campeonato.

## 2 REFERENCIAL TEÓRICO

Esta seção tem o objetivo de apresentar a base teórica necessária à compreensão deste trabalho, ainda que os temas sejam diversos e de conexão não imediata. Primeiramente, é feita uma revisão sobre experimentos em chaveamento, é apresentada a distribuição de probabilidades Beta, a inferência Bootstrap, uma introdução aos grafos, a escala hedônica e, por fim, as redes sociais, a produção e o consumo de queijos no Brasil e uma revisão sobre a música sertaneja brasileira.

### 2.1 CAMPEONATOS EM CHAVEAMENTO

Considere os sorteios por meio de chaveamento, que são os mais utilizados em competições. A cada confronto, se a equipe perdedora sai da competição e a vencedora continua a seguir o chaveamento, diz-se que o campeonato é do tipo “mata-mata”. Resende (2018) analisa os sorteios em esquema de chaveamento e a confiança dos resultados produzidos, como função do número de equipes. Nessa obra, o autor descreve alguns fatores envolvidos no contexto, como oportunidades de pontuar e seleção de atletas. Com o uso de um algoritmo criado em linguagem R, o autor define os conceitos de força e probabilidade de vitória de cada equipe em alguns tipos de confronto.

No intuito de investigar se os sorteios em chaveamento geram o mesmo número de combates para cada par de competidores, Resende (2018) fez simulações em R, e comparou com casos reais. A partir desta perspectiva, o autor conclui que os resultados de diferentes modelos são divergentes. No cenário 1 foi usado um modelo do tipo *playoffs*, enquanto no cenário dois, as equipes foram separadas em patamares diferentes de força. Os resultados apresentados nesse trabalho indicam que, para as equipes fracas, não é vantajoso o formato de mata-mata, já para as equipes fortes é interessante, pois as favorecem.

Como era de se esperar, a área esportiva figura como a aplicação mais frequente de campeonatos em chaveamento. Nesse contexto, as informações probabilísticas buscadas têm como principal objetivo as apostas em jogos esportivos. Com o aumento da popularidade das apostas em jogos esportivos online, houve o crescimento do interesse na análise e na previsão de resultados dessas competições. As previsões destes eventos geralmente são baseadas em dois tipos de informação: força (habilidade) dos competidores e probabilidade de vitória dos competidores. Leitner, Zeileis e Hornik (2010) propuseram uma forma de mapear as habilidades a partir das probabilidades de vitória (e vice-versa) por meio de simulação. Ambas as técnicas foram empregadas para a previsão dos resultados do Campeonato Europeu de Futebol 2008 (UEFA EURO 2008). O modelo de consenso forneceu previsões mais precisas de resultados dos torneios do que aqueles baseados nas habilidades atuais da equipe ou do jogador, em torneios anteriores. Esse modelo previu a final real do torneio e mostrou os efeitos devido ao sorteio do grupo. O modelo considerado continha apenas o efeito fixo de equipe, mas poderia abranger outros efeitos fixos e

aleatórios, como por exemplo pontos fortes do grupo ou vieses das casas de aposta.

Por sua vez, Gupta (2015), estudando campeonatos da liga norte americana de basquetebol universitário masculino (NCAA), apresenta um sistema de classificação para equipes usando dados de jogos da temporada anterior ao torneio. As classificações desse sistema são usadas em um modelo que usa quatro preditores para produzir previsões de chaves para cada torneio de 2009 a 2014. Esse modelo foi construído sob a restrição de que duas equipes têm probabilidade combinada de 100% de ganhar um determinado jogo. Esse artigo também realiza simulação Monte Carlo para investigar se são necessárias modificações a partir de um sistema de predição baseado em valor esperado, para fim de obter a pontuação máxima dentro de um grupo definido. Três preditores foram retidos na aplicação desse método: a proporção de vitórias contra as 10 melhores equipes, o número de aparições nos últimos 3 anos em torneios da NCAA e a presença (ou ausência) em uma conferência. O método superou a estratégia Pick the Seeds, os três métodos de Kaplan e Garstka, e o modelo logístico de Brady West para as temporadas de 2008-2009 a 2013-2014.

No futebol, Guyon (2015) estuda a Copa do mundo da FIFA, onde um sorteio de oito grupos que compõem a primeira etapa da Copa é feito com base em quatro princípios: aleatoriedade, tratabilidade, equilíbrio e separação geográfica. Guyon (2015) examinou criticamente esse procedimento, apontando uma série de falhas, como: os grupos não têm o mesmo nível competitivo, equipes mais fortes apresentam mais chances de terminar em grupos classificados como difíceis do que outras e assim a distribuição se torna desigual, pois todos os resultados possíveis não são prováveis. Para amenizar tais efeitos, o autor propõe um procedimento que consiste em construir oito urnas por nível e adicionar uma restrição logística, mantendo o princípio básico do sistema atual, para garantir maior equilíbrio, diversidade geográfica e uniforme. Sugeriu-se também sortear primeiro os continentes e depois as equipes.

Assim como nos esportes, a compreensão do processo de escolha de alimentos por consumidores também pode ser abordada pelo chaveamento. Quando consumidores escolhem um de dois alimentos (seja em uma feira-livre, em um supermercado ou em pesquisas nas redes sociais) é como se fosse um confronto do tipo “mata-mata”. Daí a importância de compreender o comportamento dos consumidores de alimentos.

De acordo com Hamilton, McIlveen e Strugnell (2000), para melhorar a dieta das pessoas, é preciso compreender de forma detalhada como é feita a escolha dos alimentos. Nesse estudo, foram investigadas as escolhas alimentares de jovens de 9 a 17 anos na Irlanda do Norte, sabendo que tais escolhas são influenciadas por fatores como sexo e idade. Os jovens desta faixa etária estão mais preocupados com o “aqui e o agora” do que com implicações futuras das más escolhas da sua alimentação.

Para essa pesquisa, os autores selecionaram cinco escolas para entender as características de grupos e indivíduos. A escolha das escolas não se deu de forma aleatória, e sim

de acordo com a localização e capacidade para incluir diversos grupos socioeconômicos. As visitas ocorreram durante os intervalos de almoço e a observação ocorreu nos refeitórios. Os assuntos a serem observados foram selecionados por meio de amostragem sistemática e os tamanhos das amostras foram calculados segundo a confiança e margem de erro desejados. Foram amostrados jovens na faixa etária de 9 a 17 anos e estes não eram avisados da pesquisa.

Como principais conclusões, Hamilton, McIlveen e Strugnell (2000) entenderam que é necessário repensar a forma de como a alimentação saudável é retratada para esse grupo de pessoas, gerando assim um desafio futuro para pais, escola e anunciantes, o que requer uma cooperação entre esses grupos. As escolhas alimentares dos adolescentes são pouco discutidas e têm sido reconhecidas como um “quebra-cabeça” completo de influências conscientes e inconscientes.

Outro assunto que tangencia a escolha de alimentos é a segurança alimentar, ou seja, a possibilidade de alimentos afetarem a saúde humana, tanto benéfica como prejudicialmente (se não seguro). Em pesquisa recente, Cerroni, Notaro e Raffaelli (2019) afirmam que 97,1% das amostras de alimentos continham resíduos de pesticidas dentro dos limites legais em 2014. Destacando-se os riscos à saúde devido à presença destes resíduos, foi usado, um experimento de escolha discreta, em um trabalho interessado em reduzir o risco de resíduos de pesticidas em maçãs. Os autores usaram um modelo de escolha binária, em que há apenas dois estados possíveis, cada um caracterizado pela probabilidade de um certo número de maçãs apresentar agrotóxicos acima dos limites legais.

A incerteza na tentativa de elucidar como as pessoas respondem a informações probabilísticas, como é percebido e avaliado como ruim ou boas notícias, motivou o trabalho de Cerroni, Notaro e Raffaelli (2019), que conta com a primeira tentativa de controlar crenças probabilísticas, atitudes de risco e preferências temporais em uma avaliação de bem-estar. Os resultados apresentados nesse trabalho, sugerem que as notícias que divergem das crenças dos entrevistados afetam suas preferências. Ou seja, os entrevistados atribuem pesos diferentes aos resultados probabilísticos quando estes divergem de suas crenças, e tornam-se mais avessos aos riscos.

Assim, de acordo com Cerroni, Notaro e Raffaelli (2019), podemos afirmar que sem controlar o efeito de boas e más notícias, podem ocorrer estimativas de bem-estar com um viés ascendente. Portanto, as crenças probabilísticas e o fornecimento de informações, afetam as preferências dos entrevistados por políticas alimentares incertas e a longo prazo.

Para modelar o problema probabilístico que se apresenta neste trabalho é necessário conhecimentos básicos sobre a distribuição Beta de probabilidades, apresentados na próxima seção.

## 2.2 A DISTRIBUIÇÃO BETA

A Distribuição Beta é uma família de distribuições de probabilidade contínuas amplamente utilizada em Probabilidade e Estatística. Essas distribuições são definidas no intervalo  $[0, 1]$  ou  $(0, 1)$  e são caracterizadas por dois parâmetros positivos, representados aqui por  $a$  e  $b$ . O valor de  $a$  aparece na sua densidade como o expoente da variável ( $X$ ), e  $b$  como o expoente de seu complemento ( $1 - X$ ), controlando assim a forma da distribuição.

Devido à sua versatilidade, a distribuição Beta tem sido aplicada em diversas áreas do conhecimento para modelar o comportamento de variáveis aleatórias limitadas a intervalos de comprimento finito. Souza *et al.* (2018) utilizam em seu trabalho modelos de regressão beta, com o objetivo de estudar a distribuição da obesidade em adultos de diferentes nações. Soares *et al.* (2019) utilizam um modelo de regressão beta com o objetivo verificar a relação entre a ocorrência da esquistossomose e os fatores socioeconômicos no Estado de Pernambuco, pois este modelo é útil para modelar taxas ou proporções.

Macedo, Albuquerque e Moralles (2017) verificaram a viabilidade econômico-financeira de um potencial parque de geração de energia eólica em diferentes localidades do Brasil considerando o risco por meio das técnicas auxiliares de simulação de Monte Carlo e a distribuição Beta.

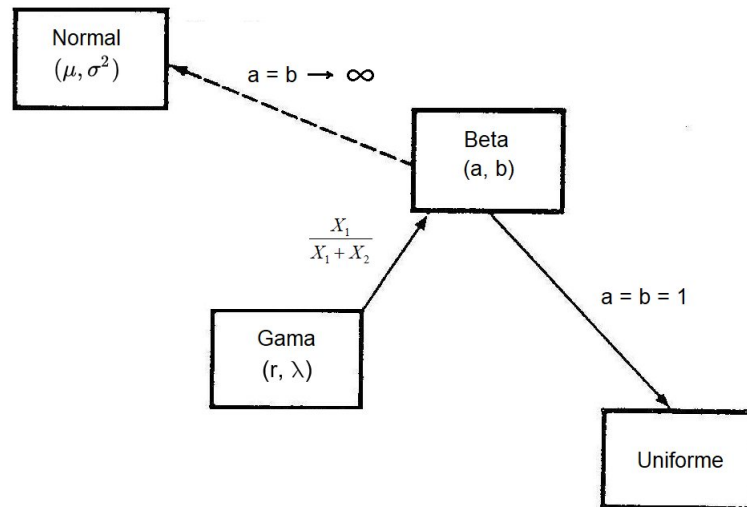
Denski e Back (2021) analisaram a aderência de diferentes distribuições de probabilidade aos dados de evapotranspiração de referência (ET0) decenal para Urussanga, SC. Para a estimativa dos valores de ET0 decenal com probabilidades de ocorrência de 5 a 95 por cento foram testadas seis funções densidade de probabilidade: Beta, Normal, Log-Normal MV, Log-Normal MM, Gama MV e Gama MM. Para os valores de ET0 decenal observou-se que as seis distribuições de frequência analisadas ajustaram-se adequadamente as frequências observadas de evapotranspiração de referência, possuindo valor de  $D_{máx}$  inferiores ao  $D_{crit}$ . No entanto, a distribuição Beta teve melhor aderência que as demais distribuições.

Na Inferência Bayesiana, a distribuição Beta desempenha um papel fundamental, pois atua como a distribuição de probabilidade a priori conjugada para várias outras distribuições, incluindo as de Bernoulli, Binomial, Binomial Negativa e Geométrica (Gelman *et al.*, 1995).

Segundo Leemis (1986), a distribuição Beta se relaciona com a distribuição Normal, quando seus parâmetros são iguais e tendem a infinito. A Beta também pode ser entendida a partir da distribuição Gama, quando tomamos uma proporção de uma na soma de duas variáveis Gama. Além disso, um caso particular da distribuição Beta é a distribuição Uniforme padrão, que surge quando seus parâmetros são iguais a 1.



Figura 1 – Relacionamento da distribuição Beta com outras distribuições.



Fonte: Adaptado de Leemis (1986).

### 2.2.1 Densidade, parâmetros e formas

A densidade de probabilidade Beta, com dois parâmetros de forma  $a$  e  $b$ , é expressa da seguintes maneiras análogas:

$$\begin{aligned} f(X|a, b) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} X^{a-1} (1-X)^{b-1} \\ &= \frac{1}{B(a, b)} X^{a-1} (1-X)^{b-1}, \end{aligned} \quad (1)$$

sendo  $0 \leq X \leq 1$ ,  $a > 0$ ,  $b > 0$ .

Os parâmetros  $a$  e  $b$  são simetricamente relacionados por:

$$f(X|a, b) = f(1-X|a, b); \quad (2)$$

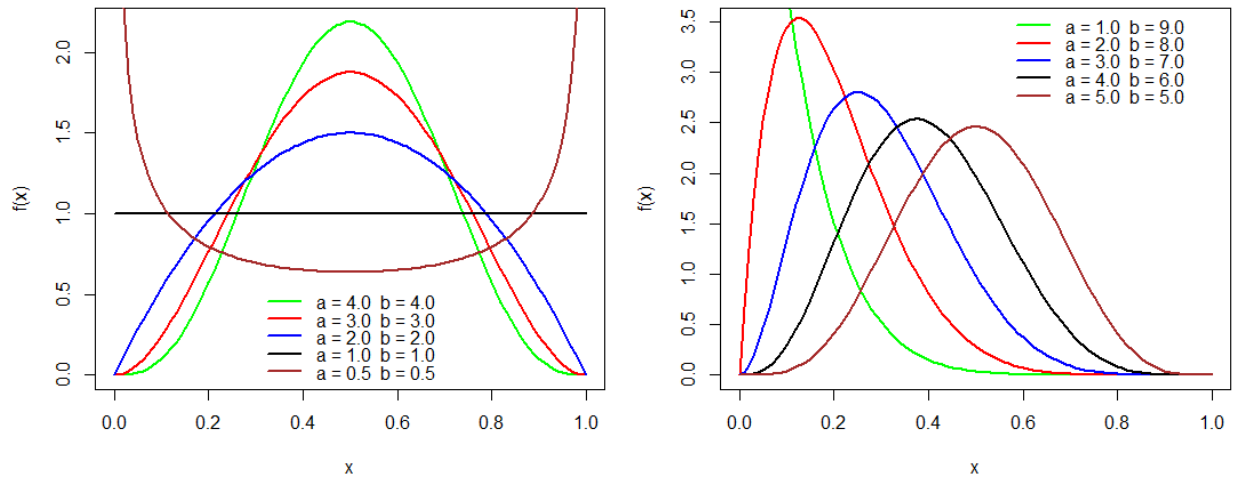
isto é, se  $X$  tem uma distribuição Beta com parâmetros  $a$  e  $b$ , então  $1-X$  tem uma distribuição Beta com parâmetros  $b$  e  $a$  (Gupta; Nadarajah, 2004).

A distribuição Beta pode apresentar uma mudança drástica em sua forma, à medida que seus parâmetros são alterados, passando de assimétrica à esquerda a assimétrica à direita, passando pela simetria.

Por exemplo, quando  $a = b$ , a distribuição é unimodal e simétrica em torno de 0,5. Observe que  $a = b = 1$  é equivalente à distribuição uniforme (0,1). A distribuição torna-se mais pontiaguda à medida que  $a$  e  $b$  aumentam (Figura 2a). Densidades Beta unimodais e enviesados. A distribuição é fortemente assimétrica à direita quando  $b > a$ , mas a distribuição fica menos assimétrica e a moda se aproxima de 0,5 à medida que  $a$  e  $b$  se aproximam um do outro. As distribuições seriam assimétricas à esquerda se os valores

de  $a$  e  $b$  fossem trocados.

Figura 2 – Formas simétricas (a) e assimétricas (b) da distribuição Beta.



Fonte: Dos autores.

Legenda: Simetria:  $a = b$ . Assimetria:  $a \leq b$  e  $a + b = 10$ .

Em toda região do espaço paramétrico em que  $a > 1$  e  $b > 1$ , a distribuição é unimodal, com moda em  $Mo = (a - 1)/(a + b - 2)$  (Mood; Graybill; Boes, 1974). Esse resultado é facilmente demonstrável. A ideia é encontrar as condições de existência para o valor de  $x$  que gera o máximo da densidade (1). Seja a sua derivada:

$$\begin{aligned}
 f'(X) &= \frac{df(X)}{dX} \\
 &= \frac{1}{B(a, b)} \frac{d}{dX} X^{a-1}(1-X)^{b-1} \\
 &= \frac{1}{B(a, b)} [(a-1)X^{a-2}(1-X)^{b-1} - X^{a-1}(b-1)(1-X)^{b-2}] \\
 &= \frac{1}{B(a, b)} X^{a-2}(1-X)^{b-2} [(a-1)(1-X) - (b-1)X]
 \end{aligned}$$

que, ao ser igualada a zero, resulta em

$$\begin{aligned}
 f'(X) &= 0 \\
 (a-1)(1-X) - (b-1)X &= 0 \\
 a-1 - X(a-1+b-1) &= 0 \\
 X &= \frac{a-1}{a+b-2}
 \end{aligned}$$

Assim, impondo as exigências que o denominador deve ser  $a+b-2 \neq 0$  e  $0 \leq X \leq 1$ , a moda existe (e tem valor único) quando  $a > 1$  e  $b > 1$ .

Prosseguindo a análise das possíveis formas da densidade Beta, a distribuição se apresenta fortemente assimétrica à direita quando  $b$  é muito maior que  $a$ , mas a distribuição fica menos assimétrica quando  $a$  e  $b$  se aproximam. (Figura 2b) As distribuições

seriam assimétricas à esquerda se os valores de  $a$  e  $b$  fossem trocados.

Quando  $a < 1$  e  $b < 1$ , a distribuição é em forma de U e assimétrica com uma antimoda em  $x = (a - 1)/(a + b - 2)$ . A distribuição fica menos assimétrica e a antimoda se aproxima de 0,5 quando  $a$  e  $b$  se aproximam. O caso em que  $a = b = 0,5$  é conhecido como distribuição arco-seno, usada na teoria da comunicação estatística e no estudo do método passeio aleatório simples. A distribuição empurra a massa do centro para as caudas a medida que  $a$  e  $b$  diminuem (Figura 2a).

Mudar os valores de  $a$  e  $b$  reverteria a direção. Ainda, se  $a > 1$ ,  $b \leq 1$  a distribuição é estritamente crescente, uma distribuição beta em forma de J tribulação sem moda ou antimoda. A distribuição torna-se mais curva à medida que  $b$  diminui. Alternar os valores de  $a$  e  $b$  produz o inverso da distribuição Beta em forma de J (Figura 2b).

## 2.2.2 Estimação dos parâmetros

As técnicas convencionais para estimar os parâmetros da distribuição Beta incluem a abordagem de máxima verossimilhança e o método dos momentos. As equações de máxima verossimilhança para a distribuição Beta não possuem uma solução direta, tornando necessário recorrer a métodos iterativos para encontrar as estimativas. Por outro lado, os estimadores obtidos pelo método dos momentos são mais simples e podem ser calculados de forma direta.

### 2.2.2.1 Estimadores de Máxima Verossimilhança

Uma técnica amplamente reconhecida para estimar parâmetros é a abordagem da máxima verossimilhança. Essa abordagem envolve a definição da função de verossimilhança para uma amostra aleatória de tamanho  $n$ ,  $\mathbf{X}^t = [X_1 \ X_2 \ \dots \ X_n]$ , retirada de uma população parametrizada por  $\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \dots \ \theta_k]$ , com densidade  $f(\mathbf{X}|\boldsymbol{\theta})$ . Dessa forma, pode-se escrever a função de verossimilhança  $L(\boldsymbol{\theta}|x_1, \dots, x_k) = \prod_{i=1}^n f(\boldsymbol{\theta}|x_i)$ .

O estimador de máxima verossimilhança (EMV) é o valor do parâmetro que torna o valor observado na amostra mais provável. Os EMVs em potencial são obtidos como soluções da equação  $\frac{\partial}{\partial \theta_i} L(\boldsymbol{\theta}|x) = 0$ ,  $i = 1, \dots, k$ . É possível confirmar que esses pontos correspondem a máximos, ao examinar a segunda derivada da função de verossimilhança para assegurar que ela seja negativa.

Frequentemente, é mais conveniente lidar com o logaritmo da função de verossimilhança, pois isso simplifica os cálculos,  $\log L(\boldsymbol{\theta}|x)$ , como derivadas de somas são mais atraentes do que derivadas de produtos (Casella; Berger, 2002).

Os estimadores de máxima verossimilhança (EMVs) são preferíveis devido à sua propriedade de consistência e eficiência assintótica. Isso significa que eles tendem, em probabilidade, a convergir para o parâmetro que estão estimando e a alcançar a menor

variância possível.

No caso particular de interesse deste trabalho, a função de verossimilhança para a distribuição Beta é dada por:

$$\begin{aligned} L(a, b|X) &= \prod_{i=1}^n \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x_i^{a-1} (1-x_i)^{b-1} \\ &= \left( \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \right)^n \prod_{i=1}^n (x_i)^{a-1} \prod_{i=1}^n (1-x_i)^{b-1} \end{aligned}$$

produzindo a log verossimilhança

$$\begin{aligned} \log L(a, b|X) &= n \log(\Gamma(a+b)) - n \log(\Gamma(a)) - n \log(\Gamma(b)) \\ &\quad + (a-1) \sum_{i=1}^n \log(x_i) + (b-1) \sum_{i=1}^n \log(1-x) \end{aligned} \quad (3)$$

Para determinar os estimadores de máxima verossimilhança (EMVs) de  $a$  e  $b$ , são calculadas as derivadas da log-verossimilhança em relação a cada parâmetro, as derivadas parciais são equacionadas a zero e, posteriormente, as equações resultantes são resolvidas.  $\hat{a}_{MV}$  e  $\hat{b}_{MV}$ :

$$\frac{\partial}{\partial a} \log(a, b|X) = \frac{n\Gamma'(a+b)}{\Gamma(a+b)} - \frac{n\Gamma'(a)}{\Gamma(a)} + \sum_{i=1}^n \log(x_i) = 0 \quad (4)$$

$$\frac{\partial}{\partial b} \log(a, b|X) = \frac{n\Gamma'(a+b)}{\Gamma(a+b)} - \frac{n\Gamma'(b)}{\Gamma(b)} + \sum_{i=1}^n \log(1-x_i) = 0 \quad (5)$$

Não há solução de forma fechada para este sistema de equações, então será resolvido para  $\hat{a}_{MV}$  e  $\hat{b}_{MV}$ , usando algum método iterativo como, por exemplo, o método de Newton-Raphson. Dessa forma, é possível calcular  $\hat{\theta}^T = [\hat{a} \ \hat{b}]$  iterativamente:

$$\hat{\theta}_{i+1} = \hat{\theta}_i - G^{-1}g \quad (6)$$

sendo  $g$  o vetor de equações normais para as quais queremos

$$g^T = [g_1 \ g_2] \quad (7)$$

com

$$g_1 = \psi(a) - \psi(a+b) - \frac{1}{n} \sum_{i=1}^n \log(x_i) \quad (8)$$

$$g_2 = \psi(b) - \psi(a+b) - \frac{1}{n} \sum_{i=1}^n \log(1-x_i) \quad (9)$$

e  $G$  é a matriz das segundas derivadas

$$G = \begin{bmatrix} \frac{\partial g_1}{\partial a} & \frac{\partial g_1}{\partial b} \\ \frac{\partial g_2}{\partial a} & \frac{\partial g_2}{\partial b} \end{bmatrix} \quad (10)$$

sendo

$$\frac{\partial g_1}{\partial a} = \psi(a) - \psi'(a+b), \quad (11)$$

$$\frac{\partial g_1}{\partial b} = \frac{\partial g_2}{\partial a} = -\psi'(a+b), \quad (12)$$

$$\frac{\partial g_2}{\partial b} = \psi'(b) - \psi'(a+b), \quad (13)$$

e  $\psi(a)$  e  $\psi'(a)$  são as funções di e tri-gama definidas como

$$\psi(a) = \frac{\Gamma'(a)}{\Gamma(a)} \quad (14)$$

e

$$\psi'(a) = \frac{\Gamma''(a)}{\Gamma(a)} - \frac{\Gamma'(a)^2}{\Gamma(a)^2} \quad (15)$$

O sistema composto pelas equações 11, 12 e 13 deve ser resolvido por algoritmos iterativos, por não possuir solução explícita, como por exemplo, o algoritmo de Newton-Raphson (Burden; Faires, 2010). Esse algoritmo converge quando as estimativas de  $a$  e  $b$  sofrem alterações inferiores a um valor tolerado a cada iteração subsequente para  $\hat{a}_{MV}$  e  $\hat{b}_{MV}$ .

### 2.2.2.2 Estimadores pelo Método dos Momentos

O método de estimadores por momentos dos parâmetros da distribuição Beta envolve a equação dos momentos da distribuição Beta com a média e a variância da amostra.

A função geradora de momento para um momento de ordem  $t$  é:

$$\begin{aligned} E(x^t) &= \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} x^t dx \\ &= 1 + \sum_{k=1}^{\infty} \left( \prod_{r=0}^{k-1} \frac{a+r}{a+b+r} \right) \frac{t^k}{k!} \\ &= \frac{\Gamma(a+b)\Gamma(a+b)}{\Gamma(a+b+t)\Gamma(a)}. \end{aligned} \quad (16)$$

O primeiro momento da distribuição Beta é, então,

$$E(X) = \frac{a}{a+b}. \quad (17)$$

O segundo momento da distribuição Beta é

$$E(X^2) = \frac{(a+1)a}{(a+b+1)(a+b)}. \quad (18)$$

Podemos então resolver a variância da distribuição Beta como

$$\begin{aligned} \text{Var}(X) &= E(X^2) - (E(X))^2 \\ \text{Var}(X) &= \frac{ab}{(a+b)(a+b+1)} \end{aligned}$$

Nosso método de estimadores de momentos são encontrados definindo a média amostral,  $\bar{X}$  e variância  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  igual a média e variância da população

$$\bar{X} = \frac{a}{a+b} \text{ e } S^2 = \frac{ab}{(a+b)^2(a+b+1)}. \quad (19)$$

Para obter estimadores de  $a$  e  $b$  resolvemos as equações 19 para  $a$  e  $b$ , resultando em:

$$\hat{a} = \bar{X} \left( \frac{\bar{X}(1-\bar{X})}{S^2} - 1 \right) \quad (20)$$

$$\hat{b} = (1-\bar{X}) \left( \frac{\bar{X}(1-\bar{X})}{S^2} - 1 \right) \quad (21)$$

Quando há restrições no conhecimento, impossibilidade ou extrema dificuldade de se obter ou assumir a distribuição de probabilidade das estatísticas amostrais de interesse, o procedimento inferencial de Bootstrap é comumente utilizado, pois se vale do esforço computacional para aproximar tais distribuições. Esse é o caso aqui, onde temos uma amostra de apenas um confronto, ou seja, amostras de tamanho um, para se estimar os parâmetros da distribuição Beta. A próxima seção traz detalhes sobre a Inferência Bootstrap.

### 2.3 INFERÊNCIA BOOTSTRAP

O bootstrap é um método computacional de reamostragem proposto por Bradley Efron (Efron, 1979), amplamente utilizado a partir da década de 90 por requerer computadores mais potentes. O método consiste em realizar amostras repetidas, um grande número de vezes, do mesmo tamanho da amostra original. Para facilitar o entendimento, imagine uma amostra original de tamanho  $n$ . A técnica consiste em realizar  $n$  sorteios, com reposição, desta amostra original, construindo novas amostras de tamanho  $n$ . O

procedimento deve ser repetido  $B$  vezes, até existirem  $B$  amostras bootstrap. Essas  $B$  amostras servem para fazer inferências e testar hipóteses acerca do parâmetro de estudo.

É uma técnica estatística utilizada para estimar a incerteza de uma estatística descritiva ou para realizar inferências sobre uma população, sem fazer suposições sobre a distribuição subjacente dos dados. O processo de reamostragem Bootstrap envolve a criação de múltiplas amostras de dados, chamadas de amostras bootstrap, por meio de seleção aleatória com reposição dos dados originais. Em outras palavras, a cada vez que uma observação é selecionada para compor uma amostra bootstrap, ela é devolvida ao conjunto de dados original antes da próxima seleção, permitindo que a mesma observação seja selecionada mais de uma vez.

A ideia subjacente ao Bootstrap é que as amostras bootstrap criadas a partir dos dados originais fornecem uma estimativa aproximada da distribuição subjacente dos dados. A partir dessas amostras, é possível calcular estatísticas descritivas, intervalos de confiança e realizar testes de hipóteses. O método bootstrap é útil, na prática, para encontrar, quando não for possível analiticamente, estimativas do viés, erro padrão, erro quadrático médio, etc. de um determinado estimador via reamostragem (Efron, 1979).

Por exemplo, suponha que você tenha um conjunto de dados com 100 observações. Usando a reamostragem Bootstrap, você pode criar várias amostras bootstrap, cada uma com 100 observações selecionadas aleatoriamente do conjunto de dados original com reposição. A partir dessas amostras bootstrap, você pode calcular estatísticas, como a média, desvio padrão, correlação, entre outras. Com base nas distribuições dessas estatísticas calculadas a partir das amostras bootstrap, você pode inferir sobre a população de onde os dados originais foram obtidos. Para Carrasco, Fabio e Santana (2022), o método bootstrap e algumas extensões pode ser usado para explicar a relação existente entre o número de medalhas obtidas pelos atletas brasileiros ao longo da história olímpica em função do número de participantes e o Produto Interno Bruto per capita. Os resultados obtidos mostraram que os métodos bootstrap utilizados para prever o número de medalhas conquistadas nas olimpíadas do Japão em 2021 é uma boa alternativa como método de estimação quando o tamanho da amostra é pequeno. Silva *et al.* (2018), em seus estudos, utilizaram o método bootstrap, que se mostrou bastante eficiente na sistematização de padronização de dados estatísticos de controle de qualidade, podendo ser utilizado para diminuir as flutuações estatísticas, como também, é ideal para construir intervalo de confiança para projetar formas de maximizar a capacidade de monitorar a qualidade de produtos que estão sendo produzidos e eliminar os defeitos.

### 2.3.1 Bootstrap Não-Paramétrico

O Bootstrap, como mencionado anteriormente, é uma abordagem estatística que envolve a reamostragem dos dados existentes para estimar a distribuição de uma estatística de interesse sem fazer suposições sobre a distribuição subjacente dos dados. Quando esse

método é aplicado a técnicas não paramétricas, significa que estamos usando a reamostragem Bootstrap para avaliar ou realizar inferências sobre estatísticas não paramétricas. Estatísticas não paramétricas são aquelas que não dependem de suposições específicas sobre a forma da distribuição dos dados.

Para Davison e Hinkley (1997), os problemas não paramétricos mais simples, obtêm-se, literalmente, amostras a partir da amostra, sendo que é razoável ser considerado que a amostra contém toda a informação sobre a população. É necessário assumir que os valores da amostra são permutáveis, ou seja, não se altera quando a ordem das observações é trocada por meio de uma permutação. E nos casos onde a distribuição da variável aleatória é desconhecida, ou não pode ser tratada de forma crítica, atribui-se a cada elemento desta amostra a mesma probabilidade  $1/n$  de ser selecionado, usando a distribuição de probabilidade para a reamostragem.

### 2.3.2 Bootstrap Paramétrico

O Bootstrap paramétrico é uma abordagem estatística que utiliza o método Bootstrap para realizar inferências ou estimar a incerteza em técnicas paramétricas. Em contraste com o Bootstrap não paramétrico, o Bootstrap paramétrico assume que os dados seguem uma distribuição paramétrica específica. Quando se aplica o Bootstrap paramétrico, assume-se que a distribuição subjacente dos dados segue um modelo paramétrico conhecido ou estimado a partir dos próprios dados. Essa distribuição paramétrica é usada para gerar as amostras bootstrap.

O bootstrap paramétrico assume uma distribuição de probabilidade para a variável aleatória amostrada, os parâmetros são desconhecidos, a distribuição de amostragem do estimador também é desconhecida, utilizam-se estimativas dos parâmetros de interesses obtidas a partir da amostra que se tem como parâmetro da distribuição de probabilidade assumida, gerando assim novas amostras. Os motivos pelos quais se utiliza um método de reamostragem é para determinar o estimador de um parâmetro de interesse e a avaliação de sua precisão.

Como o agrupamento gerado pelos confrontos analisados neste trabalho não é linear, serão utilizados grafos para apresentar o resultado do campeonato e o relacionamento entre os competidores. Por isso, a próxima seção apresenta o conceito de grafos.

## 2.4 GRAFOS

Um grafo é uma estrutura matemática que consiste em um conjunto de elementos chamados vértices, também conhecidos como nós, que estão interconectados por linhas chamadas arestas (também conhecidas como arcos). A origem da teoria dos grafos remete ao século XVIII, associada ao problema das pontes de Königsberg, antiga Prússia. Parte desta cidade, localizava-se entre duas ilhas do Rio Pregel, as quais estavam ligadas através



de sete pontes. Os habitantes da cidade desejavam passar pelas sete pontes, sem passar duas vezes pela mesma. Em 1736, Leonard Euler (1707-1783), analisou o problema trocando as áreas de terras por pontos (vértices) e as pontes por arcos (arestas). Ele provou que o problema não tinha solução recebendo hoje o nome de Teoria dos Grafos.

Para Aquino (2023), os grafos têm ganhado destaque justamente pelo potencial de aplicação a problemas típicos da sociedade atual como problemas de transporte, alocação, telecomunicações e também pelo potencial de desenvolver habilidades de coleta, sistematização e análise de dados, investigação e problematização, bem como por permitir a resolução de problemas da realidade em sala de aula. Bastos (2013) ressalta as várias situações nas quais é importante representa o inter-relacionamento entre um conjunto finito de objetos e que os grafos são estruturas de dados presentes na ciência da computação que permitem representar centenas de problemas computacionais.

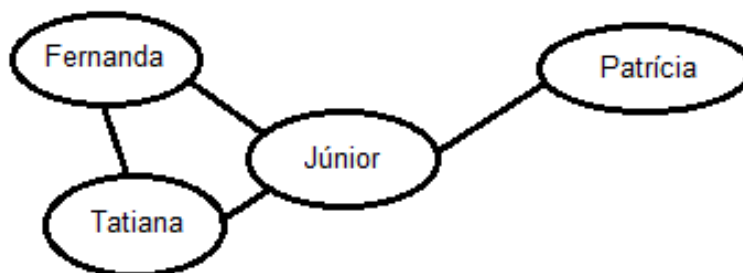
Aqui estão listados alguns exemplos de como os grafos são aplicados em diferentes áreas. Sua estrutura rica em representar relacionamentos e conexões torna-os uma ferramenta poderosa para modelagem e análise em uma ampla gama de domínios:

- a) **Redes de Computadores:** Grafos são usados para modelar redes de computadores, onde os nós representam dispositivos, como computadores ou roteadores, e as arestas representam as conexões entre eles;
- b) **Redes Sociais:** Grafos são usados para modelar relacionamentos em redes sociais, como amizades em plataformas online. Os nós representam usuários e as arestas representam conexões de amizade;
- c) **Sistemas de Recomendação:** Grafos são usados em sistemas de recomendação para identificar padrões de comportamento e relacionamentos entre usuários e itens. Isso ajuda a fazer recomendações personalizadas com base nas preferências dos usuários;
- d) **Rotas e Transporte:** Grafos são usados para modelar sistemas de transporte e encontrar rotas mais eficientes. Os nós podem representar locais, como cidades, e as arestas podem representar estradas ou conexões de transporte;
- e) **Bioinformática:** Grafos são usados para modelar redes de interações entre proteínas, genes e compostos químicos em estudos de bioinformática. Isso ajuda na compreensão de interações moleculares e no desenvolvimento de medicamentos;
- f) **Engenharia de Software:** Grafos são usados na análise e modelagem de dependências de código, fluxos de programa e estruturas de dados em engenharia de software. Isso ajuda na detecção de problemas, otimização de algoritmos e geração de código;
- g) **Logística e *Supply Chain*:** Grafos são usados para otimizar o fluxo de materiais, rotas de transporte e a cadeia de suprimentos em empresas de logística e gerenciamento de estoque.

Os grafos podem ser classificados em diferentes tipos, dependendo de suas características e propriedades. Alguns exemplos comuns incluem: grafo direcionado, grafo não direcionado, grafo ponderado, grafo cíclico e grafo acíclico. Para Maquengo (2019), os grafos também podem ser classificados em trivial, hipergrafos, grafos orientados e não orientados, grafos rotulados e não rotulados, grafos valorados e não valorados, grafo simétrico, grafo completo, grafo complementar, subgrafos, grafo bipartido, grafo planar, etc. A teoria de grafos serve de uma ferramenta importante, com a utilização de algoritmo, para o estudo de redes. A aplicação da teoria de grafos em rede distribuição de energia elétrica, por exemplo, possibilita o desenvolvimento, integração e aplicação de tecnologias que promovem a transformação da infraestrutura de energia elétrica para prover estudos de maneira eficiente e barata.

Essa representação gráfica também pode ser usada para descrever relacionamentos ou conexões entre objetos. Os grafos são uma estrutura de dados versátil e amplamente utilizada em várias áreas. Andrade, conclui que a teoria dos grafos se constitui numa ferramenta matemática que tem uma ampla aplicabilidade nas mais diversas áreas, como computação, engenharias, biologia, entre outras. A ideia é, em geral, utilizar grafos para simplificar análises e auxiliar em processos de tomada de decisão. Evidentemente, há a necessidade de se modelar o problema em questão com base na teoria dos grafos. Essa teoria pode ser aplicada em diversas áreas como física, química, engenharias, psicologia, etc.

Figura 3 – Grafo em que os vértices representam pessoas e as arestas representam amizade.



Fonte: Dos autores.

Um grafo é composto de vértices ou nós ( $V$ ) e arestas ( $A$ ). Por exemplo, o grafo  $G(V, A)$  é definido pelo par de conjuntos  $V$  (vértices) e  $A$  (arestas ou pares ordenados  $a = (v, w)$  em que  $v$  e  $w$  pertencem a  $V$ ). Se definimos que  $V = \{p | p \text{ é uma pessoa}\}$  e  $A = \{(v, w) | v \text{ é amigo de } w\}$ , esta definição representa toda uma família de grafos. Agora, especificando um pouco mais o exemplo, suponha que  $V = \{\text{Fernanda, Tatiana, Patrícia, Júnior}\}$  e  $A = \{(\text{Fernanda, Tatiana}), (\text{Fernanda, Júnior}), (\text{Tatiana, Júnior}), (\text{Patrícia, Júnior})\}$ . Gráficamente, este grafo pode ser apresentado pelo diagrama da Figura 3.

Faria, Alves e Barroso (2015) afirmam em seu trabalho que a geometria de grafo permite a resolução de problemas complexos que envolvam distâncias não euclidianas, tais como a busca de rotas mais curtas, rápidas ou otimizadas para instalação de facilidades, a partir de critérios pré-definidos. Uma das possibilidades de resolução do problema geográfico proposto neste trabalho é a do caminho mínimo. A determinação do caminho mínimo entre dois vértices de um grafo tem como objetivo minimizar a distância entre eles, sendo que a distância poderá ser (espaço, tempo, custo, etc). Os grafos podem ser representados por diversas formas, dentre as quais se destacam a por meio de matrizes e a geométrica, que facilita a compreensão, principalmente de informações espaciais.

Neste trabalho, além do agrupamento não linear ser apresentado em grafos, as estimativas do parâmetros da distribuição Beta recebem interpretação prática em termos de aceitação dos competidores por parte de quem vota nas enquetes do Instagram. Para isso, a próxima seção apresenta a escala hedônica, que será utilizada para proporcionar tal interpretação.

## 2.5 ESCALA HEDÔNICA

É chamada de escala hedônica, uma escala do tipo Likert - com 3, 5, 7 ou 9 pontos - que auxilia na avaliação de prazer ou desprazer de um indivíduo, geralmente associada a uma experiência, um estímulo ou um produto. As escalas hedônicas são compostas por categorias, nas quais os descritores verbais podem ser transformados em números. Esses números são distribuídos em intervalos uniformes assegurando assim que a distância entre os pontos consecutivos seja coerente. Lim (2011) afirma em seu trabalho que ao longo da primeira década dos anos 2000, houve um aumento significativo no interesse em torno da medição das respostas hedônicas na Psicometria e Sensometria, principalmente para avaliar os fatores envolvidos na formação de preferências, escolhas alimentares e segmentação do mercado consumidor.

A escala hedônica de 9 pontos é a mais utilizada, e sua criação se deu em 1947, no Instituto Quartermaster de Alimentos e Contêineres para as Forças Armadas, e foi motivada pela necessidade de uma escala de classificação que pudesse superar as limitações do complicado método de emparelhamento comparações (Peryam, 1950; Peryam; Pilgrim, 1957). Desde então, tem sido utilizada para uma infinidade de propósitos, como testar a preferência e aceitabilidade de alimentos e bebidas, governos, pesquisadores acadêmicos e diversas outras áreas, podem ser adaptada conforme a natureza da pesquisa ou estudo em questão. Na escala hedônica de 9 pontos, os descritores verbais variam de “1-desgostei extremamente” a “9-gostei extremamente”, havendo um ponto neutro no centro – o ponto 5 - que é descrito como “nem gostei e nem desgostei”. Portanto, trata-se de uma escala estruturada e ordinal (Tabela 1).

Para Wichchukit e O’Mahony (2015), as classificações derivadas de todas as escalas hedônicas devem dar a mesma ordem de classificação, sendo todos os outros fatores

Tabela 1 – Escalas hedônicas de 3, 5, 7 e 9 pontos.

3 pontos	5 pontos	7 pontos	9 pontos	Categorias
		7	9	Gostei extremamente
	5	6	8	Gostei muito
3			7	Gostei moderadamente
	4		6	Gostei ligeiramente
2	3	4	5	Indiferente
	2		4	Gostei ligeiramente
1		3	3	Gostei moderadamente
	1	2	2	Desgostei muito
		1	1	Desgostei extremamente

Fonte: Autores.

iguais. Essas variações da escala hedônica, podem ser adaptadas conforme a natureza da pesquisa ou estudo. Algumas variações dessa escala, com abordagens distintas, são: escala de 7 pontos, escala de 5 estrelas, escala de hedonismo sensorial, escala de aprovação-desaprovação, dentre outras. Para Santos *et al.* (2010) o processo de escolha da escala a ser adotada para avaliar atributos sensoriais não é tão simples. Aspectos como simplicidade de compreensão; representatividade acurada da percepção do provador; e influência do tipo de produto sobre a resposta dada pelo provador como expressão de sua avaliação acerca de um atributo podem ser importantes para aumentar a eficiência de uma análise sensorial. Por exemplo, as escalas hedônicas podem ser numéricas, nominais (ou verbais) ou mistas, independente do número de classes que possuir (Figura 4).

Figura 4 – Escalas hedônicas de 9 pontos mista, nominal e de 3 pontos mista.

<u>9 pontos mista</u>	<u>9 pontos nominal</u>	<u>3 pontos mista</u>
( ) 9 - Gostei extremamente	( ) Gostei extremamente	
( ) 8 - Gostei muito	( ) Gostei muito	
( ) 7 - Gostei moderadamente	( ) Gostei moderadamente	( ) 3 - Gostei
( ) 6 - Gostei ligeiramente	( ) Gostei ligeiramente	
( ) 5 - Indiferente	( ) Indiferente	( ) 2 - Indiferente
( ) 4 - Desgostei ligeiramente	( ) Desgostei ligeiramente	
( ) 3 - Desgostei moderadamente	( ) Desgostei moderadamente	( ) 1 - Desgostei
( ) 2 - Desgostei muito	( ) Desgostei muito	
( ) 1 - Desgostei extremamente	( ) Desgostei extremamente	

Fonte: Autores.

Addo-Preko, Amissah e Adjei (2023) investigaram a adequação do uso da escala hedônica de 9 pontos em testes de aceitação do consumidor em Gana, considerando diferenças culturais. Utilizando pão como exemplo, foram testadas escalas de 9, 7, 5 e 3

pontos. Independentemente do número de categorias, a ordem de preferência dos produtos permaneceu a mesma: pão de açúcar > pão de manteiga > pão de chá. A atenção dos avaliadores à tarefa afetou como as diferentes escalas foram utilizadas, sugerindo que o comprimento da escala pode não influenciar a pontuação de aceitação se os avaliadores estiverem atentos às mudanças na escala. A escolha do comprimento da escala deve ser considerada cuidadosamente em testes de aceitação do consumidor.

Para Wichchukit e O'Mahony (2015), o poder de discriminação, sensibilidade e a confiabilidade das escalas hedônicas de diferentes números de pontos de categorias verbais são base para determinar o quão relevante é uma escala para os consumidores ao pontuar o pão, por ser um produto universal. No trabalho de Godoy e Rodrigues Neto (2022), com o objetivo de realizar a análise sensorial do queijo frescal e temperado de leite de ovelha Santa Inês, para a avaliação dos tratamentos foi aplicada a análise sensorial com 80 provadores não treinados, utilizando-se uma escala hedônica de 9 pontos para os parâmetros cor, sabor, textura, aroma e aspecto geral.

Bastos *et al.* (2015) afirmam que um dos campos mais valorizados da ciência dos alimentos é a análise sensorial, na qual se avalia o produto por meio dos sentidos básicos (olfato, paladar, audição, visão e tato), com o objetivo de melhorar sua qualidade, atendendo às expectativas dos consumidores. Uma ferramenta importante nessa avaliação é a escala hedônica expressando o grau de gostar ou desgostar do consumidor em relação ao produto que está sendo analisado.

Santos *et al.* (2010), em seu trabalho para avaliar a precisão das escalas de mensuração, adotou duas abordagens: uma com base em estatísticas de dispersão; e outra examinando dados relativos à opinião dos provadores sobre qual escala eles consideram mais precisa para expressar sua percepção sensorial. Os resultados fornecem evidências de que a influência do atributo e do tipo de produto sobre a precisão das escalas em testes de aceitação não é significativo.

Vieira e Bandeira (2020), com o objetivo de avaliar a aceitabilidade de petiscos (*snacks*) elaborados com massa de queijo saborizada com páprica doce, chumichurri e açafrão com pimenta, utilizaram a escala hedônica de 9 pontos para avaliar a aceitação das amostras, e concluíram que é viável tecnicamente a produção de um petisco à base de queijo desidratado e saborizado com grande potencial para comercialização no mercado.

Apesar de muito utilizada, é importante considerar que, assim como qualquer método de pesquisa, a escala hedônica tem suas limitações. Pode haver variações na interpretação das respostas, e a subjetividade dos participantes pode influenciar os resultados. Portanto, é crucial projetar cuidadosamente a escala, garantir a clareza das instruções e, se possível, combinar as avaliações subjetivas com dados objetivos sempre que aplicável.

A escala hedônica é utilizada neste trabalho para interpretar a aceitação dos seguidores pelos competidores, ao votarem em batalhas, nas enquetes no Instagram. Por

isso, a próxima seção traz uma revisão sobre redes sociais.

## 2.6 REDES SOCIAIS

As redes sociais são canais importantes para os fluxos de informação. Por se manterem com base na confiança entre atores, levam à modificação e ampliação do conhecimento (Carneiro; Villas Boas, 2019). Hoje, as redes sociais são os principais meios de comunicação e informação da sociedade, pois apresentam um alto poder de disseminação de informação e um potencial “viral” que é proporcionado pela instantaneidade e interatividade que alcança uma infinidade de públicos. Em particular, o Instagram, do grupo Meta, apresenta novas possibilidades de abordagem da informação, pois exibem e modificam nossas formas de comportamento e consumo.

Além do consumo, outra forma de avaliar a popularidade, é a frequência de comentários nas redes sociais eletrônicas. Essa aquisição de dados é possível graças a ferramentas que permitem uma coleta automatizada, diretamente da Web, a raspagem dos dados (*Web Scraping*). Após a coleta, os dados são organizados em uma base estruturada, para posteriormente serem analisados (Singrodia; Mitra; Paul, 2019).

O Instagram é uma plataforma de mídia social amplamente reconhecida, onde os usuários podem compartilhar fotos e vídeos, aplicar filtros e fazer edições visuais antes de publicá-los. Desde seu lançamento em 2010, o Instagram alcançou uma posição de destaque, tornando-se uma das redes sociais mais populares globalmente, sendo um dos aplicativos mais utilizados. Ele facilita a conexão entre pessoas de diferentes partes do mundo, tornando-se uma rede cada vez mais interativa e subjetiva. Para Ramos e Martins (2018), a rede social Instagram constitui-se como uma textualidade contemporânea oriunda da cibercultura, onde autores amplamente letrados para tecnologia digital, apropriam-se de ferramentas que impulsionam a circulação de seus texto imagem e alcançam um público significativo de seguidores que, coexistem enquanto leitores.

Os *stories* são um recurso do Instagram de publicação de imagens, vídeos, textos, enquetes, caixas de perguntas, músicas, *hashtags*, localização, horários e temperatura local, figurinhas, GIFs, contagens regressivas e testes em tempo real, caracterizados por durar apenas 24 horas e desaparecer em seguida (Lemos; Sena, 2018).

Uma das crescentes tendências nos *stories* são as batalhas. Seguindo um esquema de chaveamento - aleatorizado por sorteio - competidores são colocados para duelar (dois a dois) utilizando-se a ferramenta “enquete”. Dessa forma, oponentes apresentados em figuras lado a lado, os seguidores daquela conta são convidados a votar no seu preferido (Ferreira; Valadares; Silva, 2023). Dessa forma, é possível captar a opinião dos seguidores, ao eleger o campeão de um campeonato, o segundo colocado, dois terceiros colocados, quatro quartos colocados, e assim por diante. O problema é que esse agrupamento gerado é fruto de apenas uma amostra e carece de procedimentos inferenciais adequados para computar a incerteza existente, compreendendo um campeonato como uma amostra de

uma população conceitual de infinitos campeonatos semelhantes.

Este trabalho, além de trazer uma proposta metodológica de como inferir sobre a igualdade dos competidores e sobre o agrupamento gerado pelo campeonato, ainda ilustra o procedimento com dois campeonatos conduzidos no Instagram: um sobre queijos e outro sobre cantores sertanejos. É por isso que as duas próximas seções versam sobre esses conteúdos.

## 2.7 PRODUÇÃO E CONSUMO DE QUEIJOS NO BRASIL

O Brasil é uma das principais economias do mundo. As perspectivas para o futuro mercado alimentar no país são bastante positivas, apesar da estimativa de crescimento da economia brasileira estar abaixo da média mundial. Para Ribatejo (2015), a percepção do consumidor brasileiro dos produtos industrializados, está associada a alimentos congelados e refeições pré-feitas, que são percebidas como a margarina, o leite condensado ou o iogurte, não são percebidos como industriais.

Os consumidores dão bastante importância à marca dos produtos que irão consumir, sendo esta a principal razão de escolha de um produto em detrimento a outro. De acordo também com o mesmo estudo, a televisão exerce um papel importante na divulgação das informações dos produtos alimentares, seguidos de médicos, nutricionistas e internet.

Com base em tais informações, e sabendo que o Brasil possui uma estrutura sócio demográfica caracterizada por grandes desigualdades sociais, é importante ressaltar que essas diferenças são cruciais na definição de perfis de consumidores e seus padrões de consumo. Assim, a condição social pode influenciar diretamente nos hábitos alimentares dos consumidores, por exemplo, podemos citar as diferenças no preço médio dos queijos pelo país o que é um fator importante para caracterizar o seu consumo.

De acordo com o Departamento de Agricultura dos Estados Unidos (USDA, 2020), aproximadamente 35% da produção de leite do Brasil é transformada em queijo. Para Arbello (2021), muitos fabricantes de queijos fortalecem seus negócios através de aquisições e investimentos em maquinários, ampliando assim seus canais de distribuições. O Brasil foi o terceiro maior consumidor de queijo em 2019, o consumo per capita foi de 5,31 Kg/habitante/ano, o que é considerado baixo.

De acordo com a Secretaria de Agricultura do Estado de Minas Gerais (SAMG, 2022), o queijo é um dos principais produtos produzidos pelo estado de Minas Gerais, gerando renda, emprego e estimulando o turismo. De acordo com o Sindicato da Indústria de Laticínios de Minas Gerais (Silemg), de todo volume produzido no Brasil, cerca de 40% deste total da produção nacional de queijo tem Minas Gerais como origem, demonstrando assim a importância da atividade para o respectivo estado. Minas Gerais possui oito regiões caracterizadas como produtoras de queijo, respaldadas por estudos que avaliam o processo de fabricação e suas características do local de origem, como a história, a

economia, a cultura, o relevo, a altitude, vegetação, o clima, dentre outros.

O consumo de lácteos no Brasil é determinado pela renda das famílias, perfil dos consumidores e seus hábitos alimentares e de consumo, bem como suas origens, crenças, culturas e ideologias (Siqueira, 2021). Em 2018, o IBGE revelou que a Muçarela é o queijo mais consumido no país (0,809kg/pessoa/ano), seguido de queijo Minas (0,565kg/pessoa/ano), Requeijão (0,285kg/pessoa/ano), Prato (0,267kg/pessoa/ano) e Parmesão (0,074kg / pessoa/ano).

Segundo Siqueira (2019), o queijo é um dos alimentos mais nutritivos que se conhece, é um concentrado lácteo constituído de proteínas, lipídeos, carboidratos, sais minerais, cálcio, fósforo e vitamina. A classificação dos queijos baseia-se nas características decorrentes do tipo de leite utilizado, do tipo de coagulação, da consistência da pasta, do teor de gordura, do tipo de casca, do tempo de cura, etc.

Segundo Siqueira (2019), as principais características de alguns dos queijos mais populares do Brasil são:

**Muçarela:** de origem italiana, era feito originalmente somente com leite de búfala, atualmente é um dos queijos mais fabricados e consumidos no Brasil, produzido com leite pasteurizado, sua massa é filada, isto é, sua massa é finamente fatiada, aquecida e as fatias misturadas até formar um bloco liso e homogêneo, com consistência firme e compacta, cor esbranquiçada e sabor levemente ácido.

**Parmesão:** é um queijo de origem italiana, bastante popular no Brasil. De baixa umidade, semigordo, massa pré-cozida e prensada. O tempo de maturação é de, no mínimo, seis meses, podendo ultrapassar os dois anos. Feito com leite de vaca cru ou pasteurizado, possui consistência dura e textura compacta, granulosa com crosta firme e lisa. Sua cor é ligeiramente amarelada e o sabor levemente picante e salgado.

**Prato:** é um queijo gordo de média umidade, massa semi-cozida e de consistência semidura. De sabor suave, levemente adocicado, sua cor é amarelo-ouro. Tempo de maturação antes de ser consumido é de 45 a 60 dias. Feito exclusivamente de leite de vaca pasteurizado, este produto se popularizou no Brasil como uma modificação dos queijos Dambo (dinamarquês) e Gouda (holandês).

**Gorgonzola:** é produzido a partir do leite de vaca, é um queijo semiduro, não prensado. Possui consistência esfarelada e sua textura é fechada, odor próprio, sabor salgado e picante. Sua cor é branco-creme, com formações características, bem distribuídas verde azuladas, devidas ao *Penicillium roqueforti*.

**Suíço:** é feito com leite de vaca, com pasta prensada cozinhada, à qual é adicionado leite puro. É redondo e tem uma casca acastanhada, passa por um processo de refinação entre 5 a 12 meses, após esse período é mergulhado em água salgada.



**Coalho:** vêm do sertão nordestino, sua fabricação passa pela etapa de coagulação com o uso do coalho. Na etapa em que a massa se torna uma coalhada, esta é cortada, coada, prensada e salgada. É resistente ao calor e por esse motivo é o preferido para churrasco, pois pode ser assado sem derreter por completo.

**Requeijão:** sua principal característica é seu formato pastoso, a fabricação é feita com leite pasteurizado, a coagulação é obtida adicionando fermento para que ocorra a coagulação. A massa é tratada em duas etapas: fracionamento e aquecimento da massa à 50°C para retirada do soro, e a segunda etapa é a fusão acrescentando sal fundente.

**Cheddar:** é produzido com leite de vaca cru, mas também pode ser produzido com leite de ovelha e búfala. O principal diferencial é o tempo de maturação que é de no mínimo 9 meses, as outras varrições vêm do animal, região, estação e tempo de maturação.

Trabalhos de raspagem do Twitter revelaram que os queijos mais comentados nessa rede social são, nessa ordem, queijo Coalho, Requeijão, Cheddar, Prato, Parmesão, Muçarela, Suíço e Gorgonzola (Soares *et al.*, 2021). No entanto a frequência de citação dos nomes dos queijos no twitter não reflete, necessariamente, a sua preferência pelo público. Nem mesmo o consumo reflete fielmente a preferência, afinal, consome-se aquele queijo que cabe no orçamento familiar.

## 2.8 A MÚSICA SERTANEJA BRASILEIRA

A música desempenha um papel fundamental na disseminação cultural e influencia várias facetas da vida diária das pessoas. Nas duas últimas décadas, o gênero musical sertanejo emergiu como um líder proeminente nos meios de comunicação e no cenário nacional é possível observar dois movimentos antagônicos: a cultura da felicidade e a cultura da sofrência. Utilizando a música como um foco de pesquisa, torna-se viável examinar o contexto histórico, social e cultural de períodos específicos. Ao explorar as narrativas presentes nas composições musicais, podemos investigar a vida cotidiana dos indivíduos como um reflexo de suas interações sociais e das tradições e crenças que permeiam essas interações. A música sertaneja evoluiu para um produto que atualmente é produzido e consumido principalmente em ambientes urbanos, em contraste com suas raízes rurais. Ela deixou de focar o estilo de vida rural e adotou predominantemente narrativas centradas em desilusões amorosas e amores não correspondidos, características próprias do jovem citadino contemporâneo.

Para Silva (2018), a música hoje conhecida popularmente como sertaneja surge como uma forma de narrar o cotidiano do indivíduo que vivia no campo, oriundo principalmente das regiões Sul, Sudeste e Centro-oeste do Brasil. A música sertaneja, como uma

forma de expressão artística autenticamente brasileira, surgiu da fusão entre as práticas de canto e dança indígenas e as melodias introduzidas pelos colonizadores portugueses. É importante destacar que o berço desse gênero musical está nos estados de São Paulo e Minas Gerais, a partir dos quais ele se disseminou para diversas outras regiões do país. Ao longo do tempo, a natureza da música caipira, originada da vivência do indivíduo caboclo e rural, sofreu uma evolução significativa, particularmente nas décadas de 1970 e 1980.

Nesse período, houve uma mudança notável no estilo das canções, com as narrativas do cotidiano perdendo sua proeminência nas composições para dar espaço a temas românticos, especialmente relacionados a histórias de amor não correspondido. Com o advento do novo estilo de música sertaneja, que passou a ser denominado como música sertaneja romântica, ocorreu uma fusão entre a balada romântica dos Estados Unidos e o formato de canto em dueto. Chitãozinho e Xororó foram os pioneiros na introdução desse ritmo inovador. O gênero musical conhecido como sertanejo universitário incorpora canções que frequentemente abordam temas como relacionamentos casuais, amores perdidos, alegria, celebrações, consumo de bens materiais e álcool. Essa tendência musical se destaca pela fusão de estilos que incluem elementos do axé, pop, rock e sertanejo tradicional. Sua origem está associada à ascensão da dupla César Menotti e Fabiano, marcando a transição do sertanejo romântico para o universitário, durante a primeira década dos anos 2000.

Silva (2021) ressalta que a organização é social marcada tradicionalmente pelo olhar masculino sobre o feminino, e as letras do *Feminejo* surgem se contrapondo a traços dos valores sociais da cultura caipira enraizados na música sertaneja atual. O *Feminejo* fala com o público feminino trazendo nas letras uma contraposição aos símbolos pop da música centrados no masculino e aos antigos valores herdados da música caipira pelo sertanejo.

Quando analisamos o panorama do consumo de música no Brasil e o comparamos com dados de outros países da América Latina em uma visão abrangente, não podemos ignorar os números impressionantes associados ao gênero musical chamado “sofrência”. Portanto, é importante examinar cuidadosamente esse fenômeno, pois ele pode nos fornecer *insights* valiosos sobre a cultura da felicidade. Brasiliense e Seixas (2020) concluem que tanto na sofrência masculina quanto no feminejo, confirmam a necessidade social contemporânea de estar sempre em estado de felicidade, negando assim a melancolia, estado normal que faz parte da vida de qualquer pessoa.

Se a presença feminina é notável em um gênero que tradicionalmente viu o sucesso dos homens como intérpretes, a participação das mulheres não necessariamente representa uma quebra das características intrínsecas da música, incluindo o machismo. Isso ocorre porque, assim como nas canções interpretadas por homens, nas vozes femininas também encontramos referências à mulher de maneira pejorativa e objetificada.

Além disso, as artistas femininas frequentemente exploram narrativas semelhantes, como a busca pelo amor de um homem, o que mantém certas convenções do gênero. Para Brasileira e Seixas (2020), a música produz significados sociais, inspirando e/ou alienando pessoas, seja como exportadora de determinado estilo de vida, seja como elemento de mudança histórico-social-cultural. Ferreira e Gonçalves (2021) ressalta que nesse cenário midiático, esses discursos musicais podem reforçar ideias hegemônicas que constituem no sujeito ouvinte/telespectador certa identificação, de acordo com as suas subjetividades e experiências vivenciais, pois pode ser a partir disso que o mesmo interpreta essas letras, o que pode vir a reforçar ainda mais suas possíveis representações sociais de gênero feminino e masculino. Para Ferreira e Gonçalves (2021), as representações sociais de gênero feminino e masculino nas músicas sertanejas analisadas, houve uma mudança e transformações de valores culturais e que apesar das conquistas da mulher é o poder do homem que ainda prevalece nos três estilos musicais, ou seja, raiz, romântico e universitário.

Sa (2019), com o objetivo de cartografar aspectos da consolidação da Rede de Música Brasileira Pop Periférica e identificar alguns de seus atores, tomando como rastros as listas dos vídeos os musicais mais vistos no Youtube Brasil entre 2012 e 2015, pois o Youtube é a principal plataforma digital de consumo musical e um dos sites favoritos entre usuários de redes sociais no Brasil, ressaltando que o número de visualizações dos vídeos, alcançam cifras de 100 milhões. Para Sa (2019), as redes sociais, e sobretudo o Youtube, são agentes fundamentais na articulação dessa rede constituída por gêneros musicais que antes circulavam por circuitos distintos e que agora se irrigam através desta plataforma, utilizando as noções de rede, ator, rastros etc.

### 3 MATERIAL E MÉTODOS

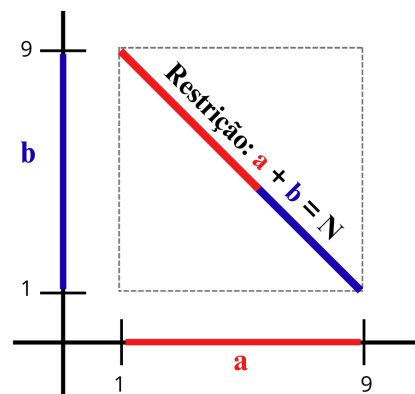
Esta seção traz a descrição da proposta metodológica deste trabalho e duas aplicações em dados reais, sendo elas, uma batalha de queijos e uma batalha de sertanejos. O objetivo da primeira subseção é introduzir os conceitos de força relativa de um competidor, força restrita e sua relação com os parâmetros da distribuição Beta, e esclarecer as razões por trás da restrição das forças.

#### 3.1 METODOLOGIA PROPOSTA

Entende-se por *força* de um competidor alguma métrica proporcional a sua probabilidade de vencer um confronto. No entanto, na modelagem proposta neste trabalho, a *força* considerada para um competidor não é absoluta, mas relativa. A força é dependente de seu oponente, pois é razoável que sua probabilidade de vencer dependa de quem ele está enfrentando.

Além disso, este trabalho traz o conceito de *força restrita*, que é a força de um competidor, restrita a um intervalo fechado, por exemplo, o intervalo  $[1, 9]$ , como em uma escala hedônica de 9 pontos. Na modelagem proposta, a força de cada um dos dois oponentes em uma batalha é entendida como um dos dois parâmetros da distribuição Beta( $a, b$ ). Em uma distribuição Beta( $a, b$ ) regular, temos que  $0 < a, b < \infty$ , mas aqui, restringimos o espaço paramétrico para  $1 \leq a, b \leq 9$  (Figura 5).

Figura 5 – Restrição do espaço paramétrico.



Fonte: Autores.

Além disso, faz sentido estipular um vínculo entre os parâmetros (entre as forças dos oponentes). Afinal, quando se diz que um oponente é mais forte que o outro, queremos dizer que um parâmetro tem o valor maior e o outro tem valor menor. Quando dois oponentes são similares, têm a mesma força ou são igualmente bons, é natural que o valor de seus parâmetros sejam iguais, ou ao menos parecidos. Desse raciocínio emerge o

vínculo de fixar a soma das forças (dos parâmetros):

$$a + b = N \quad (22)$$

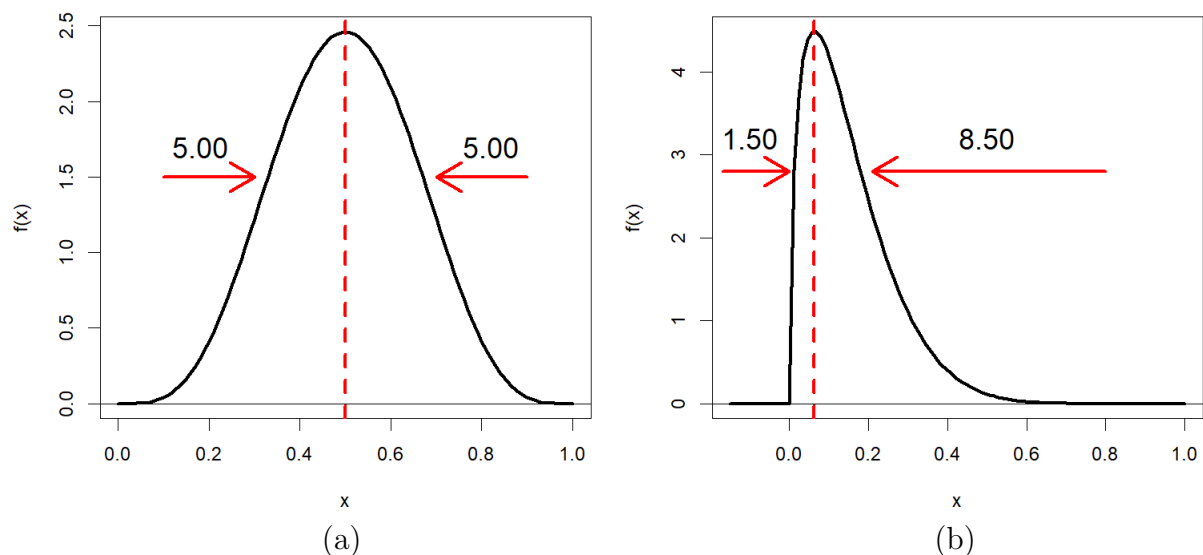
em que  $N \in \mathbb{R}_+$  é uma constante arbitrária, e  $1 \leq a, b \leq N - 1$ . Neste trabalho, assume-se  $N = 10$  para que  $1 \leq a, b \leq 9$  (Figura 5). Vale notar que ao vincular os dois parâmetros, a densidade Beta se torna uniparamétrica.

A escolha de forças serem restritas entre 1 e 9, repousa em duas razões:

1) **Metafórica.** Os valores de 1 a 9 fazem alusão à escala hedônica de 9 pontos, tipo de escala *Likert* mais utilizado em estudos de aceitação com consumidores, que varia de “desgostei extremamente” a “gostei extremamente” (Wichchukit; O’Mahony, 2015).

2) **Estatística.** Valores dos parâmetros  $(a, b)$  nessa faixa produzem distribuições Beta com as formas simétrica, assimétrica à direita e assimétrica à esquerda (Figura 2b), evitando formatos constantes, J e U para a distribuição. A forma da distribuição tem direta relação com a força do competidor. De forma lúdica, é como se, na distribuição Beta, cada oponente empurrasse a distribuição em uma direção. Quem for mais forte, consegue empurrar mais a moda (maior densidade de probabilidade). A Figura 6 traz dois exemplos.

Figura 6 – Densidades Beta em confrontos com forças similares (a) e diferentes (b).



Fonte: Autores.

A subfigura 6a representa o confronto entre dois competidores com forças iguais, ou seja, seus parâmetros são  $a = b = 5$ , tornando a Beta ajustada simétrica. Na escala hedônica, é como se o consumidor fosse “Indiferente” quanto à preferência, ou seja, nenhum dos dois competidores é preferido.

Já na subfigura 6b, pode-se notar uma disparidade entre os dois oponentes. O oponente “da esquerda”, possuidor do parâmetro  $a = 1,5$  tem força muito menor que o oponente da direita, cuja força é  $b = 8,5$ . Por essa razão, a moda da distribuição é deslocada para a esquerda. É como se os consumidores “gostassem extremamente” do oponente da direita, enquanto “desgostassem extremamente” do oponente da esquerda.

Além disso, este trabalho utiliza uma ligeira modificação do método dos momentos, ao igualar o primeiro momento amostral à moda da distribuição Beta. Vale notar que a proporção gerada em um confronto (aqui denotada por  $y$ ) representa uma amostra de tamanho  $n = 1$  e é, também, a média amostral.

$$\frac{a - 1}{a + b - 2} = y \quad (23)$$

em que  $y$  é a proporção resultante do confronto, e  $a$  e  $b$  são os parâmetros da distribuição Beta original.

A necessidade dessa modificação se deve ao fato de a Beta poder assumir formatos muito assimétricos. Nessas situações, a moda é menos influenciada pelos valores da cauda, captando fielmente o acúmulo da densidade de probabilidade. Em outras palavras, a ideia é forçar que o pico da densidade de probabilidade acompanhe o resultado observado, reforçando que, aquilo que aconteceu é o mais provável de acontecer.

A partir das restrições (22) e (23), obtêm-se os estimadores  $\hat{a}$  e  $\hat{b}$ .

$$\begin{cases} a + b = N \\ \frac{a - 1}{a + b - 2} = y \end{cases} \Rightarrow \begin{cases} \hat{a} = y(N - 2) + 1 \\ \hat{b} = N - 1 - y(N - 2) \end{cases} \quad (24)$$

em que  $N$  é uma constante arbitrária;  $y$  é a proporção resultante do confronto<sup>1</sup>; e  $a$  e  $b$  são os parâmetros da distribuição Beta original.

Como este trabalho considerou a constante  $N$  igual a 10, os estimadores pontuais das equações (24) se resumem a

$$\begin{aligned} \hat{a} &= 8y + 1 \\ \hat{b} &= 9 - 8y \end{aligned} \quad (25)$$

Em seguida, as estimativas intervalares de  $a$  e  $b$  são obtidas por meio de reamostragem *Bootstrap* paramétrico. A ideia central é gerar milhares de observações de uma distribuição Beta parametrizada por  $\hat{a}$  e  $\hat{b}$ , e em seguida obter a distribuição empírica dos estimadores pontuais. Daí, as estimativas são ordenadas, e os quantis  $\alpha/2$  e  $1 - \alpha/2$

---

Neste trabalho, foi considerada a maior das duas proporções, sem perda de generalidade. Por exemplo, se os resultados de um confronto foram 35% e 65%, considerou-se  $y = 0,65$ .

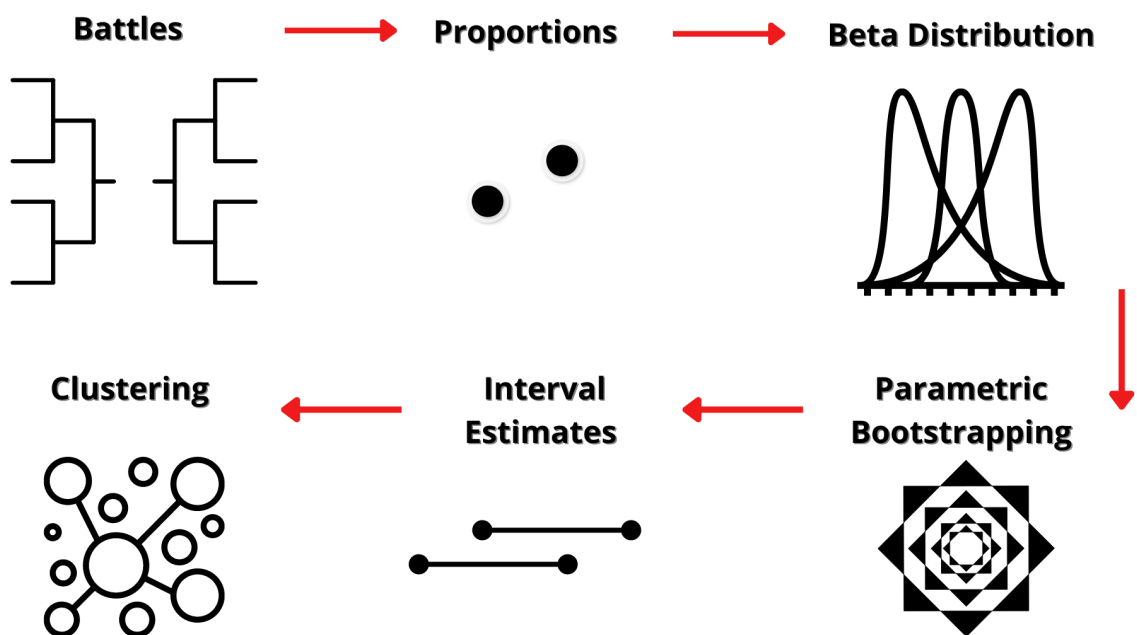
formam um intervalo de  $100(1 - \alpha)\%$  de confiança de Bootstrap para  $a$  e  $b$ :

$$IC_{(1-\alpha)\%}(a) = [q_{\alpha/2}(a); q_{1-\alpha/2}(a)] \quad \text{e} \quad IC_{(1-\alpha)\%}(b) = [q_{\alpha/2}(b); q_{1-\alpha/2}(b)].$$

Então, é feito um teste a partir da comparação de intervalos de confiança. Ou seja, os intervalos para  $a$  e  $b$  são comparados e, se houver interseção, não se pode afirmar que os parâmetros diferem a um nível  $\alpha$  de significância.

O procedimento descrito até aqui deve ser aplicado ao resultado em cada batalha. Ao fim do campeonato, tem-se a informação de todos os pares de competidores que se enfrentaram, e em quais casos a proporção exibida pode ser considerada significativa. Um resumo das etapas do método proposto está apresentado na Figura 7, onde ele pode ser melhor entendido e sumariado.

Figura 7 – Resumo do procedimento metodológico proposto.











Fonte: Autores.

Por fim, a melhor apresentação desses resultados deve ser discutida em trabalhos futuros. Neste trabalho, são trazidas três formas de apresentar os resultados: i. Uma representação do chaveamento seguido pelo campeonato, que enfatiza o fluxo mas não as diferenças significativas; ii. Um grafo com competidores nos vértices e resultado do teste nas arestas, que enfatiza quais competidores se enfrentaram e quais resultados foram significativos, mas não mostra o fluxo do campeonato; iii. Um grafo modificado, que além da significância dos relacionamentos, enfatiza a classificação final do campeonato.

### 3.2 BATALHA DOS QUEIJOS

Foi feita uma aplicação da metodologia proposta em uma batalha de queijos. Nesta aplicação, foram confrontados os oito queijos mais citados no Twitter, segundo o trabalho de Soares *et al.* (2021): queijo Coalho, Requeijão, Cheddar, Prato, Parmesão, Muçarela, Suíço e Gorgonzola. As fotos utilizadas para identificar os queijos foram adaptadas de páginas da internet e suas fontes, devidamente citadas, constam na Tabela 2.

Tabela 2 – Foto utilizada, nome e fonte dos oito queijos utilizados neste trabalho.

	Queijo	Foto adaptada de
	Cheddar	Nicolini <a href="https://shorturl.at/mBFKX">https://shorturl.at/mBFKX</a>
	Coalho	Supermercados Cidade Canção <a href="https://shorturl.at/efr06">https://shorturl.at/efr06</a>
	Gorgonzola	Supermercados Carone <a href="https://shorturl.at/EJR06">https://shorturl.at/EJR06</a>
	Muçarela	Hortifrute <a href="https://shorturl.at/dfuD4">https://shorturl.at/dfuD4</a>
	Parmesão	De La Croix <a href="https://shorturl.at/glsIP">https://shorturl.at/glsIP</a>
	Prato	Sitemercado <a href="https://shorturl.at/hmHPS">https://shorturl.at/hmHPS</a>
	Requeijão	UOL <a href="https://shorturl.at/uBHT3">https://shorturl.at/uBHT3</a>
	Suíço	Tricurioso <a href="https://shorturl.at/kqwy5">https://shorturl.at/kqwy5</a>

Fonte: Autores.

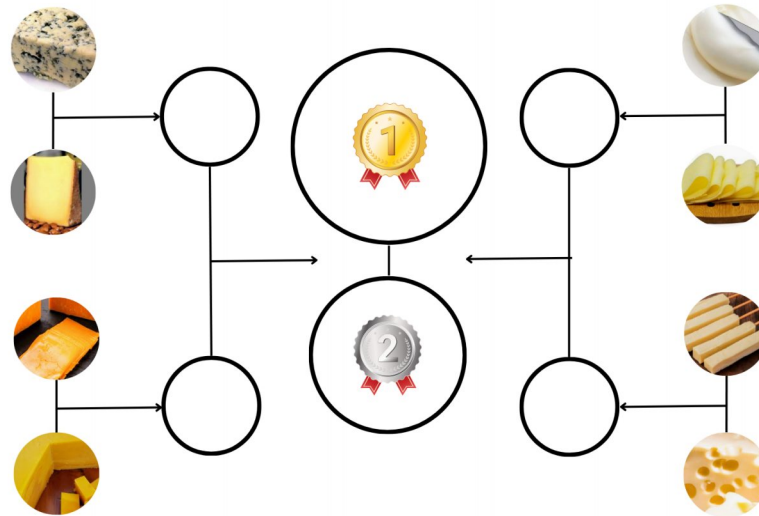
Para avaliar a preferência dos referidos queijos, foram formados pares para confrontos do tipo mata-mata, de modo inteiramente ao acaso, seguindo um chaveamento de tamanho  $2^k$ , com  $k = 3$  (Figura 8). Dessa forma, houve 3 fases de confrontos (quartas de final, semifinal e final), com  $2^3 = 8$  competidores.

As batalhas foram conduzidas nos *stories* do Instagram, no dias 13, 14 e 15 de agosto de 2022, na conta @eric.barba (Barba, 2022), utilizando a ferramenta *enquete* com duas classes, uma para o nome de cada queijo do embate.

As proporções obtidas nas batalhas - que ficaram 24 horas no ar - foram anotadas, e o número de votantes variou de 60 a 80, por batalha. Essas proporções foram usadas



Figura 8 – Sorteio dos confrontos para as chaves do campeonato dos queijos mais populares do Brasil no Twitter.



Fonte: Autores.

na estimação pontual dos parâmetros (Equações 25) que, por sua vez, abasteceram a estimação intervalar Bootstrap paramétrico, com 10.000 repetições. A cada repetição Bootstrap é gerado valores de  $a$  e  $b$ . No final se terá 10000 valores de  $a$  e 10000 valores de  $b$ . Ordenando esses valores e pegando os quantis se constrói o intervalo de confiança para cada parâmetro. De posse das estimativas intervalares de 95% de confiança, os testes foram feitos verificando se havia interseção.

Por fim, os resultados foram organizados em três dispositivos gráficos: fluxo do chaveamento, grafo e grafo modificado com intuito de enfatizar o ranqueamento. Essas formas de apresentação visual dos resultados podem ser escolhidas pelo usuário para atingir sua necessidade. Todas as análises e gráficos foram feitos no software R v.4.3.0 (R Core Team, 2023).

### 3.3 BATALHA DOS SERTANEJOS

Foi feita uma segunda aplicação da metodologia proposta, desta vez, com uma batalha de sertanejos. Nesta aplicação foram considerados os 16 artistas sertanejos que figuravam da lista “Top 50 BRAZIL” (Spotify, 2023), músicas mais tocadas no Spotify Brasil no dia 25 de março de 2023. Esses artistas estão apresentados na Tabela 3, em ordem alfabética.

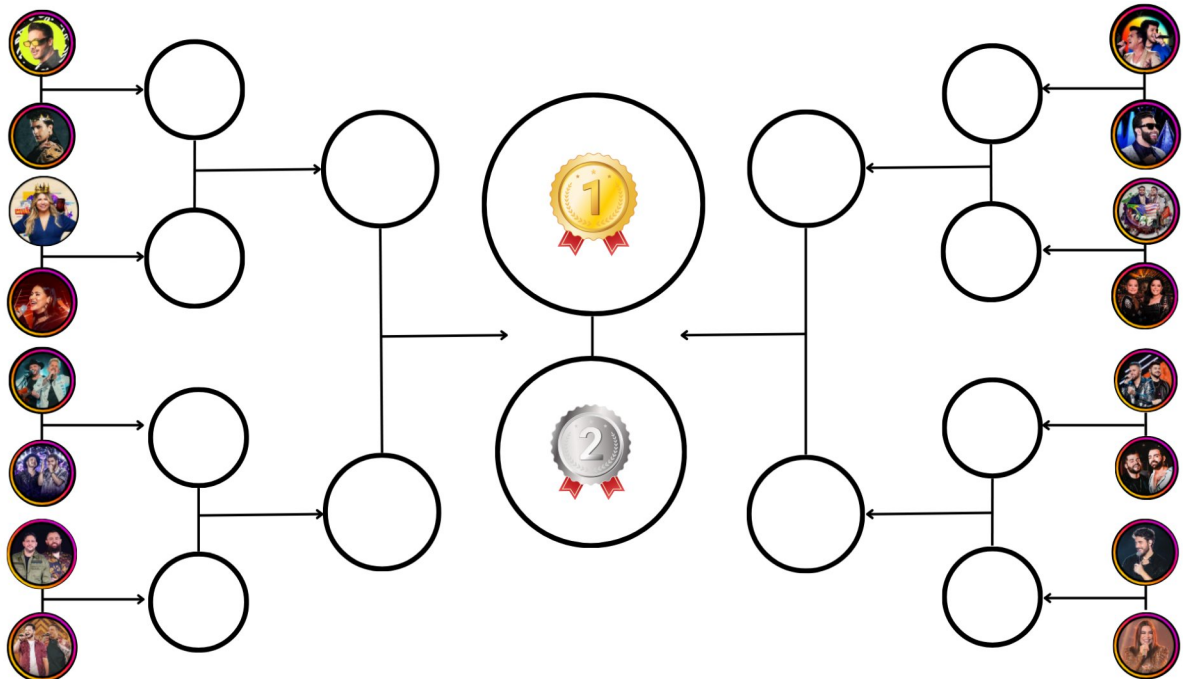
Um chaveamento duplo (com duas chaves) foi sorteado por meio da função `sample()` do software R (R Core Team, 2023), para abrigar os  $2^4 = 16$  artistas participantes desta pesquisa (Figura 9). As batalhas foram postadas diariamente (duas por dia) entre os dias 10 e 19 de abril de 2023. Foram usados os *stories* da conta @eric.barba (Barba, 2022) do Instagram.

Tabela 3 – Dezesesseis artistas sertanejos que figuravam na lista das cinquenta músicas mais tocadas no Brasil no dia 25 de março de 2023, segundo o Spotify.

Artista			Artista		
	Felipe e Rodrigo	FR		Luan Santana	LS
	Guilherme e Benuto	GB		Maiara e Maraísa	MM
	Gustavo Lima	GL		Marília Mendonça	MaM
	Gustavo Miotto	GM		Matheus e Kauan	MK
	Henrique e Juliano	HJ		Naiara Azevedo	NA
	Hugo e Guilherme	HG		Simone Mendes	SM
	Israel e Rodolfo	IR		Wesley Safadão	WS
	Jorge e Mateus	JM		Zé Neto e Cristiano	ZC

Fonte: Autores.

Figura 9 – Sorteio dos confrontos para as chaves do campeonato dos artistas sertanejos.

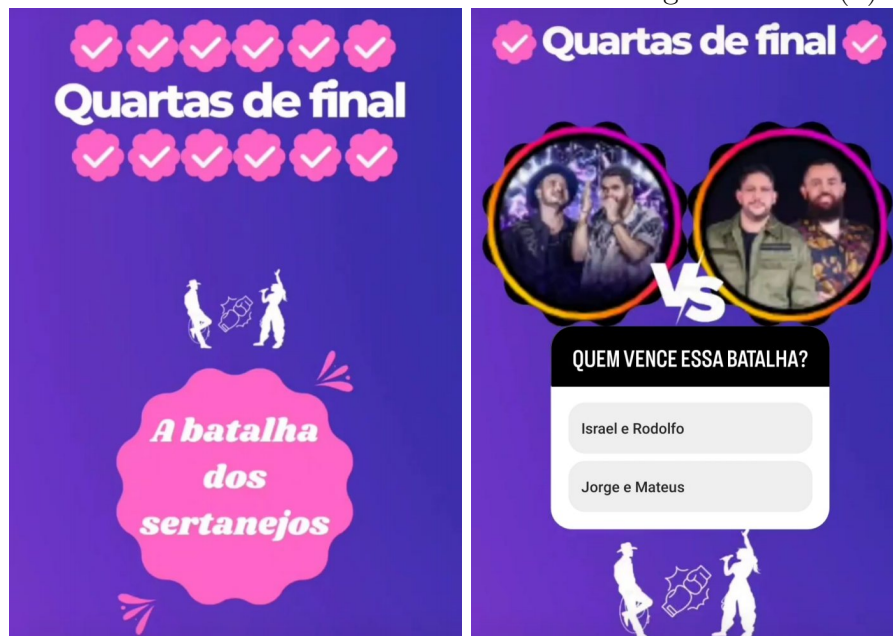


Fonte: Autores.

Foi estabelecido o mesmo leiaute para garantir o padrão de cada batalha: as fotos dos perfis dos dois artistas combatentes acrescida da ferramenta *enquete* com duas classes, uma para o nome de cada cantor (Figura 10b). As fotos dos dezesseis sertanejos combatentes foram retiradas do perfil oficial de cada artista no Instagram, data do confronto.

Antes das batalhas de cada rodada (oitavas de final, quartas de final, semifinal e final) eram postados *stories* especiais chamando a atenção dos seguidores, o que os convidava e estimulava a votar (Figura 10a). Foram usadas músicas sertanejas de artistas que não estavam participando do campeonato para ambientar os *stories* e os artistas em batalha eram sempre marcados na publicação, para caso quisessem votar ou compartilhar.

Figura 10 – *Prints* dos *stories* anunciando o início das quartas de final (a) e a batalha entre Israel e Rodolfo e Jorge e Mateus (b).



Fonte: Autores.

Cerca de 100 seguidores votaram em cada batalha. As frequências absolutas e percentuais foram computadas automaticamente pelo Instagram, e tabuladas em planilha eletrônica. Apenas a maior das duas proporções (sem perda de generalidade) foi usada, a cada batalha. Por exemplo, se uma batalha terminar em 0,6 contra 0,4, é usada a proporção 0,6. Essa proporção é utilizada para estimar os parâmetros de uma distribuição Beta, por meio dos estimadores da equação (25).

De posse das estimativas das forças dos competidores (parâmetros  $a$  e  $b$ , uma distribuição Beta foi abastecida, para cada batalha. Dessa distribuição foram sorteadas 10.000 amostras de tamanho  $n = 1$ , simulando 10.000 confrontos virtuais entre os dois sertanejos. Dessa forma um procedimento Bootstrap paramétrico é configurado.

As 10.000 observações amostradas foram então ordenadas, e os quantis de 2,5% e 97,5% ( $q_{0,025}$  e  $q_{0,975}$ ) determinados para compor o intervalo de 95% de confiança de Bootstrap. Por fim, os intervalos dos dois competidores eram comparados. Quando havia interseção, não se afirmava existir evidência de diferença significativa entre os competidores. Quando não havia interseção, afirmava-se haver evidência de diferença entre os competidores.

$$IC_{95\%}(a) = [q_{0,025}(a); q_{0,975}(a)] \quad \text{e} \quad IC_{95\%}(b) = [q_{0,025}(b); q_{0,975}(b)]$$

Por fim, o resultado final do campeonato foi organizado em três dispositivos gráficos. No gráfico do chaveamento é possível enfatizar o fluxo do campeonato. No grafo é possível enfatizar o relacionamento entre os competidores. Já no grafo modificado é possível apresentar as classificações finais dos sertanejos, quais se enfrentaram e quais enfrentamentos resultaram em vitória significativa.

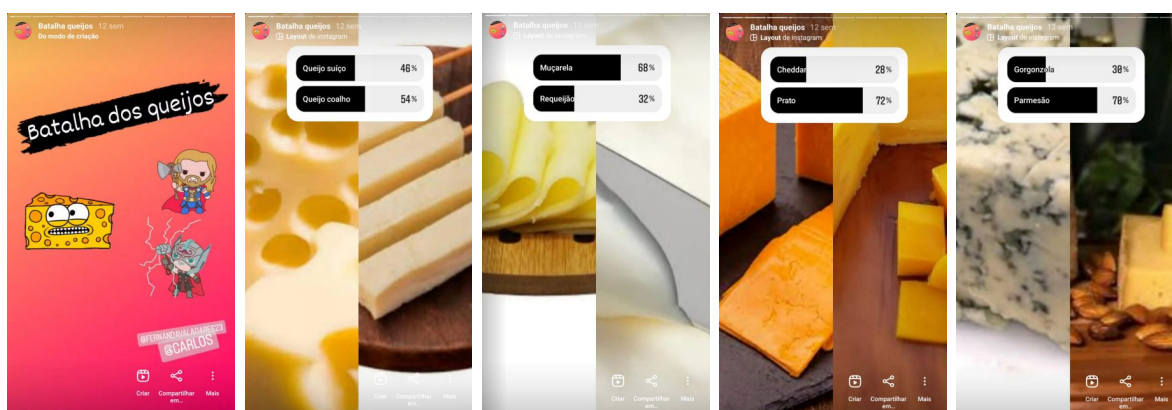
## 4 RESULTADOS E DISCUSSÃO

Esta seção traz os resultados da aplicação da metodologia proposta em duas massas de dados reais: a batalha dos queijos e a batalha dos sertanejos.

### 4.1 BATALHA DOS QUEIJOS

Seguindo o sorteio do chaveamento apresentado na Figura 8, as batalhas foram instituídas no Instagram, em uma conta pública, convidando seus seguidores a votar. A Figura 11 destaca os *stories* de capa e de cada um dos quatro confrontos das quartas de final, quais sejam: Suíço vs Coalho; Cheddar vs Prato; Gorgonzola vs Parmesão; e Muçarela vs Requeijão. Neles, é possível ver a frequência percentual de votos recebidos por cada um dos queijos competidores. As frequências absolutas de votos também puderam ser coletadas após 24 horas de votação. O total de participantes de cada confronto variou entre 60 e 80 pessoas.

Figura 11 – *Prints* dos *stories* utilizados nas quartas de final.

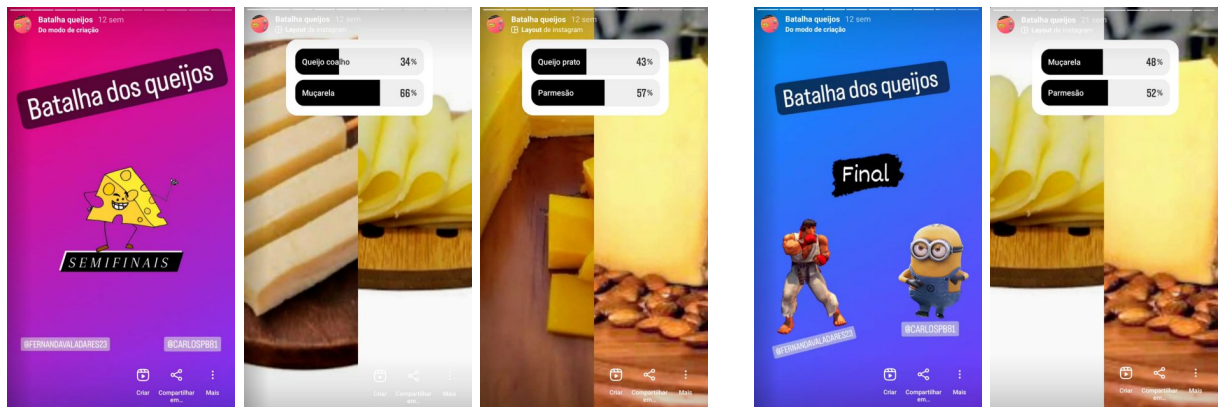


Fonte: Dos autores.

Os vencedores de cada confronto passaram para a fase seguinte: as semifinais. O queijo Coalho venceu o queijo Suíço, e foi enfrentar a Muçarela na fase semifinal. O queijo Prato venceu o Cheddar, e seguiu para enfrentar o Parmesão na semifinal. Da mesma forma, passadas as 24 horas preconizadas pelo *story*, foram coletadas as frequências absolutas e percentuais de seguidores que votaram (escolheram) cada queijo competidor. As proporções de votantes em cada competidor podem ser vistas na Figura 12, para a fase semifinal.

A Figura 12 também apresenta o *story* de capa e o resultado da batalha final, protagonizada por muçarela e parmesão. Ela mostra ambos percentuais muito próximos a 50%, o que nos leva a pensar que o queijo Parmesão pode ter vencido por mero acaso. Ou seja, se esse enfrentamento fosse repetido infinitas vezes, o queijo Parmesão poderia ter a mesma probabilidade de vencer e perder. Aliás, todos os embates revelaram proporções relativamente próximas e precisavam ser estimadas por uma estatística intervalar.

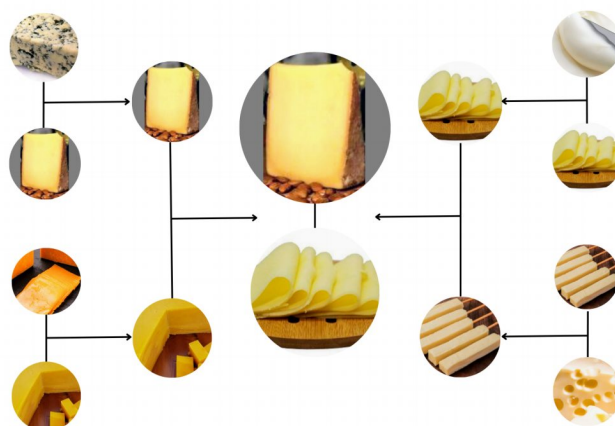
Figura 12 – Prints dos stories utilizados nas rodadas semifinal e final.



Fonte: Dos autores.

O fluxo do campeonato, ou seja, os confrontos e seus vencedores, desde as quartas de final até o resultado da final, pode ser visto na Figura 13. Na parte mais externa, destacam-se os 4 confrontos das quartas de final: Gorgonzola vs Parmesão, Cheddar vs queijo Prato, Muçarela vs Requeijão, e queijo Suíço vs queijo Coalho. A partir dos resultados dos confrontos iniciais, formaram-se os dois embates da semifinal (Parmesão vs Prato e Muçarela vs Coalho) e a grande final (Parmesão vs Muçarela). Nessa representação, pode-se ver que o Parmesão ocupa a primeira colocação, a Muçarela em segundo, Prato e Coalho em terceiro e Gorgonzola, Cheddar, Suíço e Requeijão em quarto. Note que houve 1 primeiro colocado, 1 segundo, 2 terceiros e 4 quartos. Uma limitação do chaveamento é sempre resultar em  $2^{k-1}$  grupos,  $k \in \mathbb{N}$ , sendo que todos contém um potência de 2 elementos:  $2^0$ ,  $2^0$ ,  $2^1$ ,  $2^2$ , ...,  $2^{k-1}$ .

Figura 13 – Fluxo do campeonato dos queijos.



Fonte: Dos autores.

No entanto, resta a dúvida, os vencedores passaram de fase por serem mais fortes ou por puro acaso? Para responder esta pergunta, foi aplicada a metodologia proposta neste trabalho. Após a coleta, os dados foram organizados em planilha eletrônica, formando

uma base de dados estruturada, para possibilitar a análise descritiva e inferencial dos dados. As proporções observadas em cada confronto abasteceram os estimadores pontuais dos parâmetros da distribuição Beta (equação 25), e suas estimativas - as forças restritas - foram obtidas. O valor das forças restritas foram interpretados na escala hedônica de 9 pontos, categorizando assim a opinião do consumidores votantes sobre cada queijo dentro do embate específico. A Tabela 4 apresenta, para todas as batalhas, as proporções observadas, as forças restritas (estimativas dos parâmetros  $a$  e  $b$ ) e classe da escala hedônica.

Tabela 4 – Proporções observadas em todos os combates, suas respectivas forças restritas (FR) e categoria de aceitação na escala hedônica de 9 pontos.

Fase	Batalha	Queijo	Proporção	FR	Aceitação
Quartas de final	1	Coalho	0,5455	5,364a	Indiferente
		Suíço	0,4545	4,636a	Indiferente
	2	Muçarela	0,6774	6,419a	Gostei levemente
		Requeijão	0,3236	3,581a	Desgostei levemente
	3	Prato	0,5873	5,698a	Gostei levemente
		Cheddar	0,4127	4,302a	Desgostei levemente
	4	Parmesão	0,7065	6,652a	Gostei moderadamente
		Gorgonzola	0,2935	3,348a	Desgostei moderadamente
Semi final	5	Muçarela	0,6582	6,266a	Gostei levemente
		Coalho	0,3418	3,734a	Desgostei levemente
	6	Parmesão	0,5641	5,513a	Gostei levemente
		Prato	0,4359	4,487a	Desgostei levemente
Final	7	Parmesão	0,5294	5,235a	Indiferente
		Muçarela	0,4706	4,765a	Indiferente

Fonte: Autores.

Legenda: Letras iguais indicam médias estatisticamente iguais, segundo o teste proposto.

Segundo o chaveamento tradicional, em primeiro lugar ficou o queijo Parmesão (52,9%); em segundo lugar, o queijo Muçarela (47,1%); em terceiro lugar, os queijos Prato (43,6%) e Coalho (34,2%); e em quarto lugar, os queijos Suíço (45,5%), Cheddar (41,3%), Requeijão (32,3%) e Gorgonzola (29,4%). Em comparação aos resultados encontrado no Twitter por Soares *et al.* (2021), pode-se notar que os resultados divergiram. O mesmo acontece quando se compara com os dados de consumo.

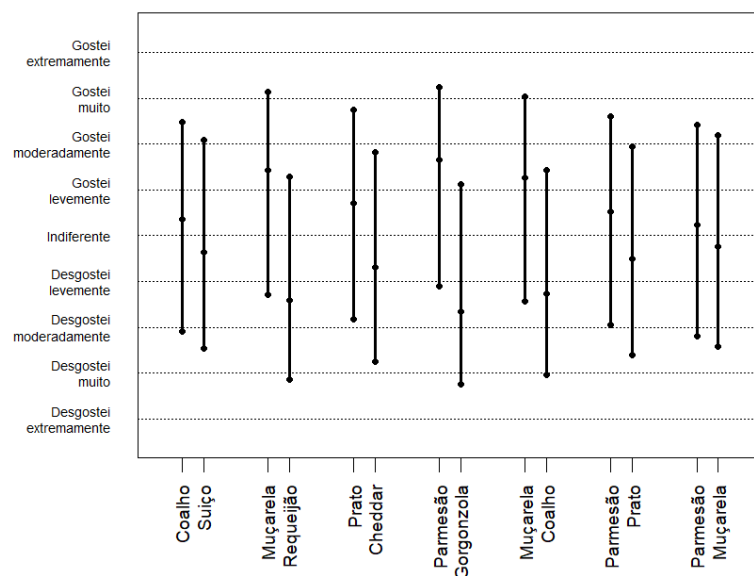
Enquanto este trabalho encontrou que os queijos mais preferidos por brasileiros são o Parmesão e a Muçarela, o trabalho de Soares *et al.* (2021) encontrou que os queijos mais comentados no Twitter são o queijo Coalho e o Requeijão. Neste trabalho, o queijo coalho se posicionou no terceiro grupo e o Requeijão, no quarto. As mais explícitas coincidências foram encontradas no queijo Gorgonzola e Prato. O queijo Gorgonzola se mostrou o menos comentado e o menos preferido, enquanto que o queijo Prato recebeu classificação

mediana nos dois tipos de estudo.

Neste contexto, são apresentados alguns preços médios obtidos em uma pesquisa na internet, em Reais por quilograma, ordem crescente. Foram priorizados os queijos alvo deste estudo. Observa-se que a grande variação pode influenciar fortemente o consumo, sendo fator agravante na escolha do queijo pelo consumidor: Muçarela: R\$30,29; Cheddar: R\$32,99; Prato: R\$33,60; Coalho: R\$76,59; Requeijão: R\$81,60; Gorgonzola: R\$83,20; Parmesão: R\$119,98; Suíço R\$90,00 a R\$120,00. Vale notar que os queijos mais preferidos, segundo este trabalho, não são necessariamente os mais baratos ou os mais caros, sugerindo uma real lembrança sensorial do consumo, e não necessariamente de uma compra.

No entanto, a estimação intervalar das forças retrizes foi fundamental. Para isso, foi aplicada a metodologia proposta neste trabalho, ou seja, as estimativas  $\hat{a}$  e  $\hat{b}$  abasteceram um processo de reamostragem bootstrap paramétrico na distribuição Beta. Na distribuição empírica das estimativas ordenadas, foram tomados os quantis de 2,5% e 97,5% para cada batalha, para obter os intervalos de 95% de confiança. Esses intervalos podem ser vistos na Figura 14. Ainda nessa figura, pode-se notar que todos os pares de intervalos possuem interseção. Dessa forma, o teste feito a 5% de significância, indica que não há razões afirmar que nenhum dos queijos difere de seus oponentes. Dessa forma, a ordenação encontrada ao final do campeonato pode ter sido encontrada por puro acaso, e os queijos serem igualmente aceitos pelos consumidores.

Figura 14 – Intervalos de 95% de confiança de bootstrap para a força restrita dos oponentes de cada batalha.



Fonte: Dos autores.

Enquanto este trabalho focou na preferência, o trabalho de Soares *et al.* (2021) averiguou a popularidade, por meio da frequência de menções. Uma terceira abordagem

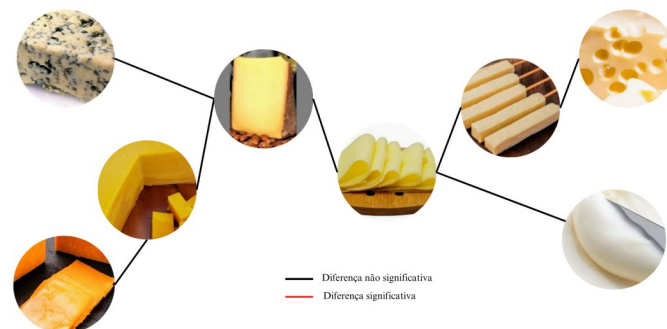


seria averiguar o consumo, o que poderia resultar em uma classificação ainda diferente. O queijo mais consumido do Brasil é a Muçarela, nas suas mais diversas formas (*Pizza cheese*, nozinho, trancinha, *sticks*, versões defumadas etc). Segundo a Associação Brasileira das Indústrias de Queijo (ABIQ, 2022), a mussarela detém 30% da produção nacional, o queijo prato, 20%, e o requeijão, 7,5%. Neste estudo, a Muçarela se apresenta como a segunda mais preferida, mas é pouco mencionada (popular) no Twitter.

Uma segunda forma de apresentar os resultados deste estudo se encontra na Figura 15, em um grafo que enfatiza o relacionamento entre queijos, mas não suas colocações no campeonato. Nesse grafo, os vértices representam os queijos e as arestas ligam os que se enfrentaram. Além disso, um sistema de cores apresenta em cor preta os queijos estatisticamente iguais e a cor vermelha os estatisticamente diferentes. Note que nenhum deles pode ser dito diferente.

Observa-se ainda que a Muçarela enfrentou três oponentes (Parmesão, Coalho e Requeijão), mas nunca foi comparada com o Cheddar, por exemplo. Já o Gorgonzola só foi comparado ao Parmesão, mas todos são estatisticamente iguais a 5% de significância.

Figura 15 – Grafo do resultado do campeonato dos queijos.

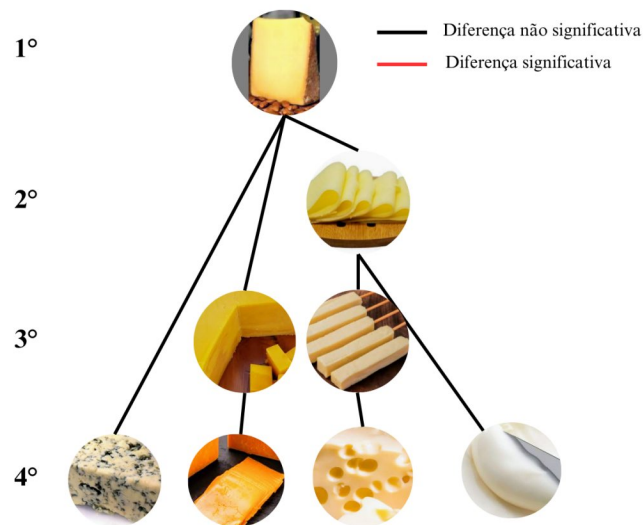


Fonte: Dos autores.

Para tentar sanar essa deficiência, foi proposto um grafo modificado (Figura 16) que, feito a partir do grafo original (Figura 15), enfatiza o ranqueamento final do campeonato. Observa-se que este é meramente um rearranjo do grafo original, de uma forma conveniente.

A representação da Figura 16 não mostra o fluxo do campeonato, mas é uma das formas mais completas de representar os resultados, uma vez que é possível ver quem se enfrentou (pelas arestas), quais diferenças são significativas (pela cor das arestas) e o ranqueamento final.

Figura 16 – Grafo modificado do resultado do campeonato dos queijos.



Fonte: Dos autores.

O consumo de queijos é um fenômeno complexo que envolve renda, oferta, procura, disponibilidade, aspectos culturais e regionais, preço, propaganda, aspectos nutricionais e calóricos, dentre outros. Nesse sentido, trabalhos futuros devem investigar mais profundamente o consumo, preferência e popularidades dos queijos no Brasil.

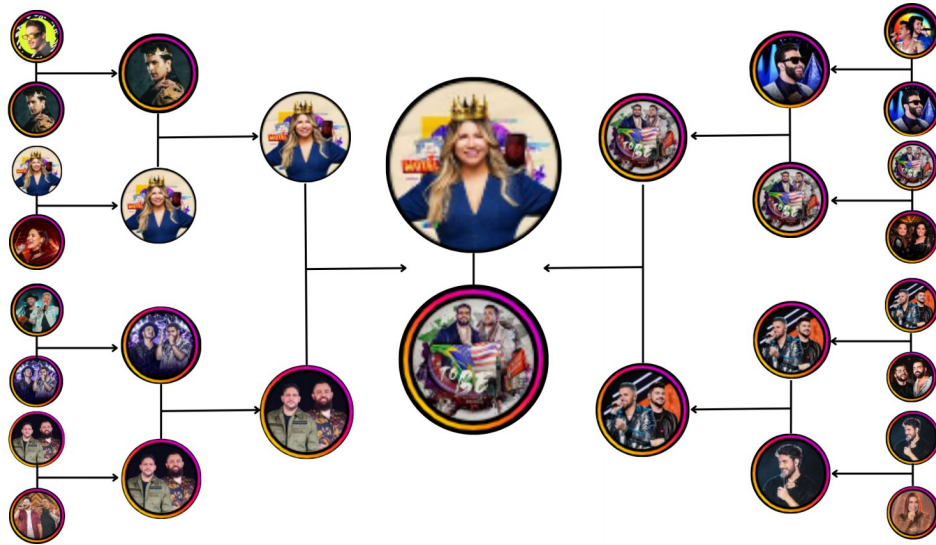
#### 4.2 BATALHA DOS SERTANEJOS

Nesta aplicação, os 16 cantores sertanejos mais populares do Brasil, segundo o ranking, se enfrentaram em um campeonato de chaveamento duplo, com 4 fases (oitavas de final, quartas de final, semifinal e final) e cujos confrontos iniciais foram sorteados conforme mostrado na Figura 9.

A Figura 17 representa o fluxo do campeonato, ela traz as duas chaves, cada batalha feita, e o progresso dos vencedores, até a fase final. Apesar de não mostrar as diferenças significativas, essa forma de apresentar os resultados enfatiza o fluxo e o ranqueamento e que a campeã foi Marília Mendonça, seguida de Henrique e Juliano, e Jorge e Mateus e Zé Neto e Cristiano.

Para inferir sobre a diferença entre as forças dos competidores, foi aplicada a metodologia proposta neste trabalho. Primeiramente, as proporções obtidas dos confrontos nos *stories* foram utilizadas para estimar os parâmetros da distribuição Beta (forças restritas) e sua interpretação na escala hedônica foi feita, mostrando a opinião dos consumidores de música sertaneja, quando cada uma das batalhas acontece. A Tabela 5 traz essas informações para todas as 15 batalhas do campeonato.

Figura 17 – Fluxo do campeonato dos sertanejos.



Fonte: Dos autores.

A batalha que exibiu a maior disparidade entre competidores foi Marília Mendonça ( $y = 0,9297$ ) versus Simone Mendes ( $y = 0,0703$ ). Nesse contexto, é compreensível supor que existem diferenças substanciais entre essas porcentagens. No entanto, quando os percentuais são semelhantes, aproximando-se de 50 por cento, surge a incerteza quanto à igualdade estatística dos competidores ou se o vencedor mereceu realmente a vitória.

No embate entre Luan Santana ( $y = 0,5172$ ) Wesley Safadão ( $y = 0,4828$ ), é possível questionar se há uma diferença significativa. Nesse caso, a vitória de Luan Santana pode ter sido puramente aleatória. Em outras palavras, se esse confronto fosse repetido inúmeras vezes, ele venceria em metade das ocasiões, enquanto o outro prevaleceria na outra metade.

Para ilustrar o uso da distribuição Beta para modelar uma batalha, veja a Figura 18. Como descrito na seção de Material e Métodos, os estimadores dos parâmetros  $a$  e  $b$  são apresentados na equação (24) e derivam da imposição de que o valor observado da proporção deve ser igual à moda da distribuição Beta desejada. Dessa forma, o ponto de densidade de probabilidade máxima da Beta “acompanha” o valor observado. A Figura 18 traz dois exemplos.

A Figura 18(a) representa o confronto entre Wesley Safadão e Luan Santana, pelas oitavas de final. Observe que o valor observado da maior proporção ( $y = 0,5172$ ) é bem central, ou seja, próximo de 0,5, tornando a Beta ajustada quase simétrica. Por consequência, as forças restritas dos dois oponentes (estimativas de  $a$  e  $b$ ) também são parecidas: 5,14 (LS) e 4,86 (WS).

Já na Figura 18(b), pode-se notar uma disparidade entre as duas oponentes. Marília Mendonça venceu Simone Mendes com uma proporção de votos observada  $y = 0,9297$ , gerando uma densidade Beta ajustada fortemente assimétrica à esquerda. Em

Tabela 5 – Proporções observadas em todos os combates, suas respectivas forças restritas (FR) e categoria de aceitação na escala hedônica de 9 pontos.

Fase	Batalha	Sertanejo	Proporção	FR	Aceitação
Oitavas de final	1	WS	0,4828	4,862a	Indiferente
		LS	0,5172	5,138a	Indiferente
	2	MaM	0,9297	8,438a	Gostei muito
		SM	0,0703	1,562b	Desgostei muito
	3	FR	0,0814	1,651b	Desgostei muito
		IR	0,9186	8,349a	Gostei muito
	4	JM	0,9111	8,289a	Gostei muito
		MK	0,0889	1,711b	Desgostei muito
	5	HG	0,3261	6,391a	Gostei levemente
		GL	0,6739	3,609a	Desgostei levemente
	6	HJ	0,6667	6,334a	Gostei levemente
		MM	0,3333	3,666a	Desgostei levemente
	7	ZC	0,5116	5,093a	Indiferente
		GB	0,4884	4,907a	Indiferente
	8	NA	0,3647	3,918a	Desgostei levemente
		GM	0,6353	6,082a	Gostei levemente
Quartas de final	9	LS	0,1186	1,949b	Desgostei muito
		MaM	0,8814	8,051a	Gostei muito
	10	IR	0,0847	1,678b	Desgostei muito
		JM	0,9153	8,322a	Gostei muito
	11	GL	0,4320	4,456a	Desgostei levemente
		HJ	0,5680	5,544a	Gostei levemente
	12	ZC	0,8430	7,744a	Gostei muito
		GM	0,1570	2,256a	Desgostei muito
Semi final	13	MaM	0,6364	6,091a	Gostei levemente
		JM	0,3636	3,909a	Desgostei levemente
	14	HJ	0,7231	6,785a	Gostei moderadamente
		ZC	0,2769	3,215a	Desgostei moderadamente
Final	15	MaM	0,7040	6,632a	Gostei moderadamente
		HJ	0,2960	3,368a	Desgostei moderadamente

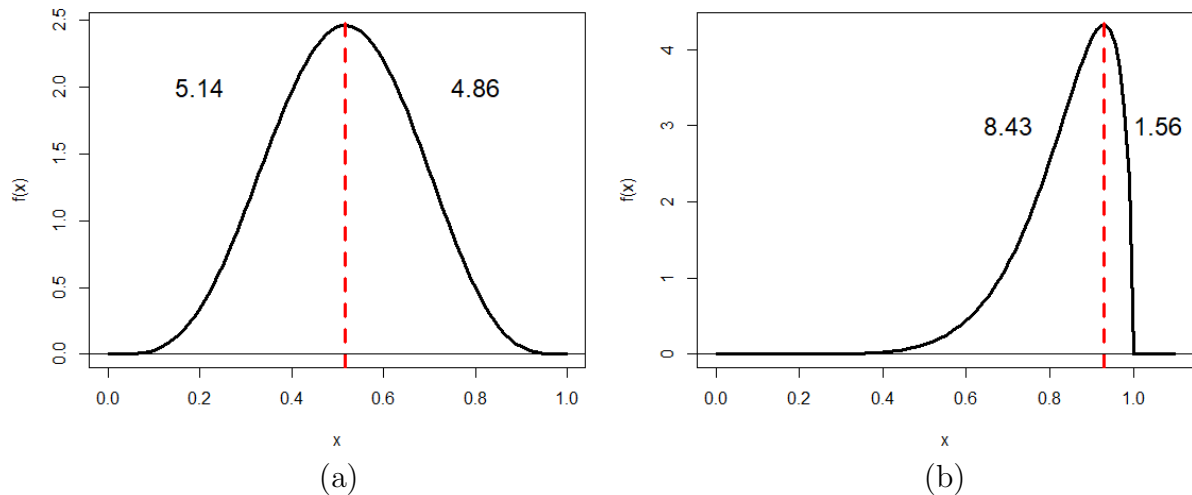
Fonte: Autores.

Legenda: Letras iguais indicam médias estatisticamente iguais segundo o teste proposto.

decorrência disso, as forças restritas estimadas para as duas competidoras também se mostraram díspares: 8,43 (MaM) e 1,56 (SM).

Em seguida, 10.000 reamostragens de Bootstrap foram feitas da distribuição Beta parametrizada por  $\hat{a}$  e  $\hat{b}$ , novamente os parâmetros  $a$  e  $b$  foram estimados, e foram cons-

Figura 18 – Exemplos de densidades Beta ajustadas a um confronto com forças restritas similares (a) e diferentes (b).



Fonte: Autores.

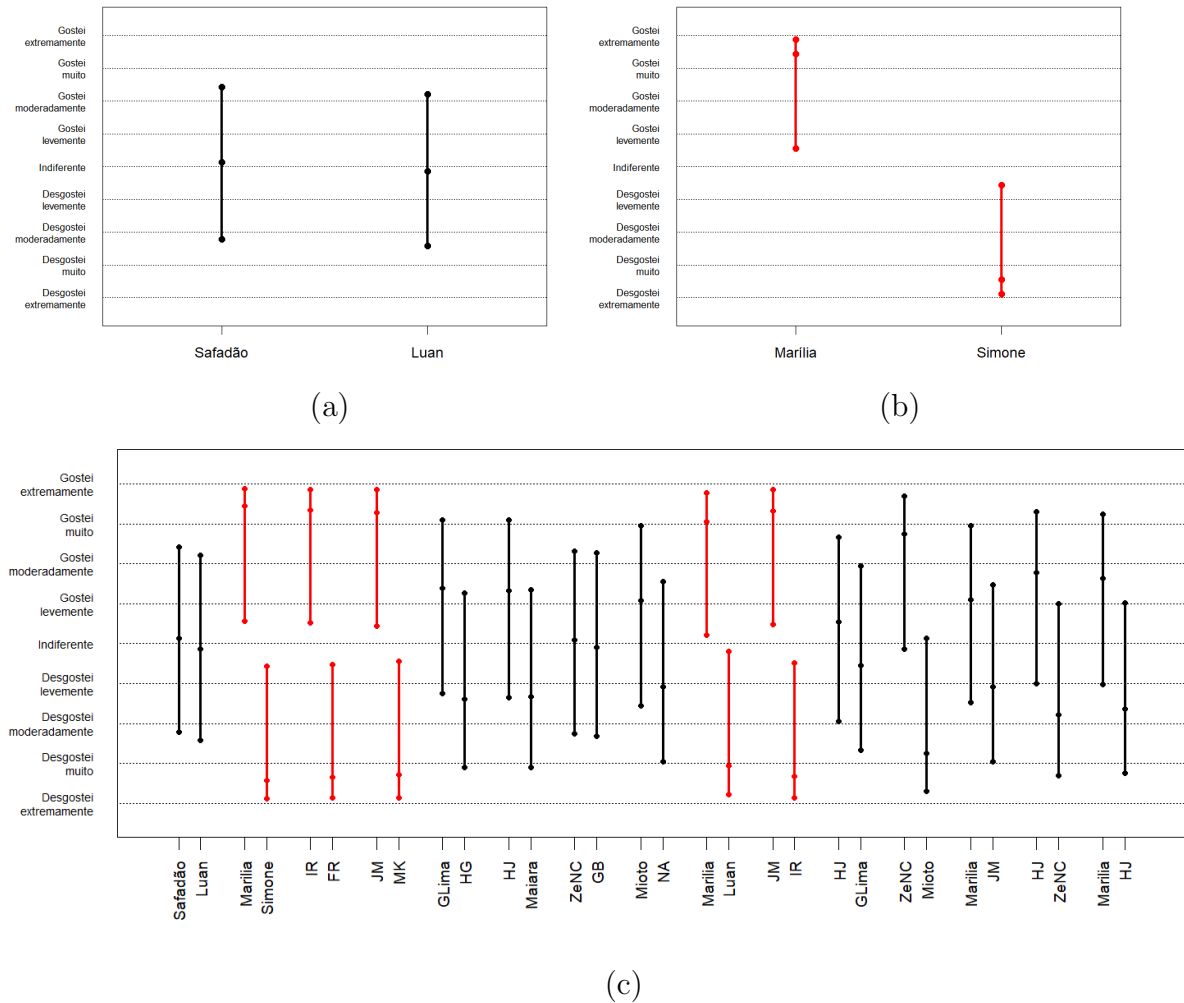
truídas distribuições empíricas de Bootstrap para  $\hat{a}$  e  $\hat{b}$ , de onde foram tomados os quantis de interesse para construir intervalos de 95% de confiança para os parâmetros. Por fim, a interseção entre os intervalos de confiança foi analisada para fazer o teste de igualdade dos parâmetros ( $H_0 : a = b$ ), a 5% de significância.

Os intervalos de confiança de Bootstrap gerados para o confronto entre Wesley Safadão e Luan Santana estão exemplificados na Figura 19a. Nele, é possível verificar a interseção entre os intervalos, e portanto, foram feitos na cor preta. Já na Figura 19b, vemos os intervalos de confiança de Bootstrap para a batalha entre Marília Mendonça e Simone Mendes. Nesse caso, é possível ver que não há interseção entre os intervalos, e por isso foram feitos na cor vermelha.

Em ambos os casos, os gráficos foram feitos utilizando-se como eixo Y a escala hedônica, que tem papel central na proposta deste trabalho. Note que os seguidores parecem ser indiferentes entre Wesley Safadão e Luan Santana, mas gostam extremamente da Marília Mendonça, enquanto, em comparação a ela, desgostam muito de Simone Mendes. Portanto, fica evidente que a interpretação da força restrita proposta neste trabalho, é relativa, e não absoluta. Em outras palavras, não significa que os seguidores desgostam da Simone Mendes. Quer dizer que eles desgostam da Simone Mendes *quando comparada* à Marília Mendonça.

De forma geral, a Figura 19c mostra as comparações entre os intervalos de confiança de Bootstrap para todas as batalhas do campeonato, mantendo o padrão de assinalar em vermelho as diferenças significativas, e em preto as não significativas. Pode-se notar que apenas 5 das 15 batalhas se mostraram significativas, o que indica que, se o campeonato fosse repetido, várias posições poderiam se inverter.

Figura 19 – Exemplos de comparações entre intervalos de 95% de confiança para as forças restritas entre os pares de competidores.

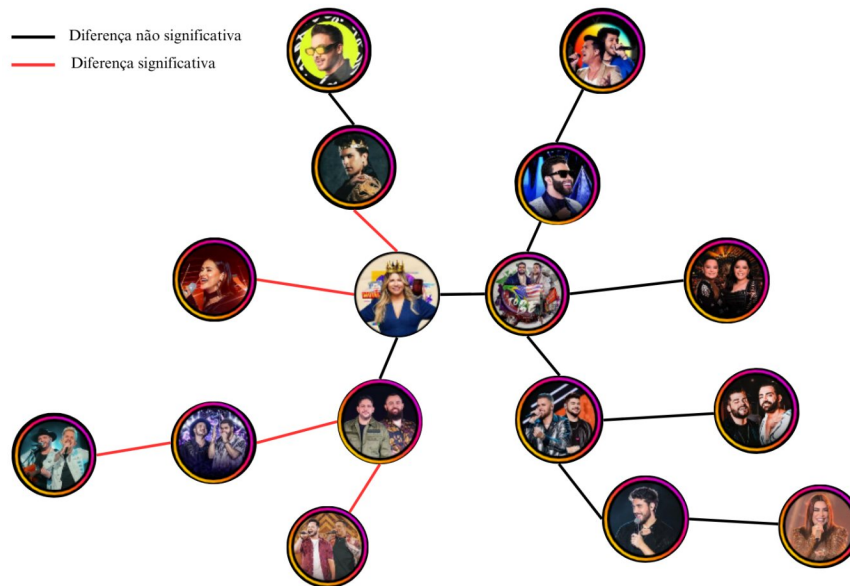


Fonte: Dos autores.

Como não houve o enfrentamento de todos contra todos participantes, não é possível fazer um agrupamento tradicional dos competidores, mostrando forças restritas médias seguidas de letras, como fazem os testes de comparações múltiplas de médias. Por isso, este trabalho traz três opções de representação dos resultados: o fluxo do campeonato (Figura 17), um grafo (Figura 20) e um grafo modificado (Figura 21).

Na Figura 20 é possível dar ênfase ao relacionamento entre os competidores. Cada vértice representa um artista sertanejo e cada aresta representa uma batalha. Quem se enfrentou possui uma aresta. Se a aresta for preta, indica que não há evidências de haver diferença entre a aceitação dos cantores. Se ela for vermelha, indica que há diferença significativa entre sertanejos. Nessa representação não se vê a evolução do campeonato nem a classificação final.

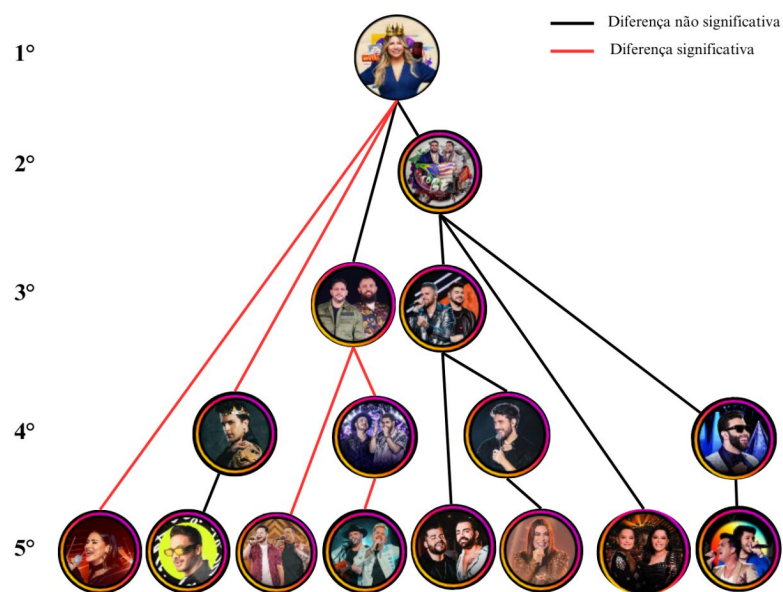
Figura 20 – Grafo com os artistas sertanejos, após o campeonato.



Fonte: Dos autores.

Sendo assim, apresenta-se o grafo modificado (Figura 21), onde podem-se observar três aspectos. Primeiro, a classificação gerada (normalmente) pelo campeonato em chaveamento duplo, que são os colocados de 1° a 5° lugar. Nota-se que o número de competidores em cada posição segue a regra de formação  $2^0, 2^0, 2^1, 2^2, 2^3 \dots$ . Para além dessa informação, as arestas informam quais sertanejos se enfrentaram, e quais combates resultaram em diferença significativa (aresta em cor vermelha).

Figura 21 – Grafo modificado do resultado do campeonato dos sertanejos.



Fonte: Dos autores.

## 5 CONSIDERAÇÕES FINAIS

Mesmo que a tendência de realizar batalhas nos *stories* do Instagram por meio de um campeonato em chaveamento perca popularidade, o tema tratado neste trabalho é importante para a modelagem de qualquer campeonato nesse formato. Campeonatos de futebol, basquete, tênis, basebol, e inúmeros outros esportes, bem como escolas de samba, concurso de miss ou qualquer outro assunto, estão sujeitos ao mesmo problema: como modelar a incerteza existente nos enfrentamentos? Isso, é claro, compreendendo um campeonato como uma amostra de uma população conceitual de infinitos campeonatos semelhantes.

Este trabalho traz uma proposta de procedimento inferencial capaz de afirmar sobre a igualdade ou a diferença entre competidores com significância fixada, baseada no uso da distribuição Beta e reamostragem Bootstrap paramétrico.

Embora outras abordagens metodológicas sejam possíveis de serem pensadas, a proposta deste trabalho se mostrou promissora e tem potencial de ser estudada, analisada e explorada em trabalhos futuros, sob diferentes pontos de vista, como as taxas de erro tipo I e tipo II praticadas.

A modelagem pela distribuição Beta permite estimar a força de um competidor e interpreta-la como a aceitação dos consumidores, frente a seu oponente. Vale dizer que não faz sentido o cálculo da força média do competidor, uma vez que ela não tem caráter absoluto, mas é relativa a seu adversário.

Na aplicação batalha dos queijos, foi detectado que o agrupamento dos queijos mais populares (mais citados no Twitter) não condiz com os mais bem aceitos pela população ou com os mais produzidos no país. No entanto, por não haver diferença significativa entre eles, a ordenação observada pode ter ocorrido por puro acaso.

Na aplicação batalha dos sertanejos, a cantora Marília Mendonça venceu o campeonato. No entanto, foi estatisticamente igual a Henrique e Juliano (que ficaram em segundo lugar) e a Jorge e Mateus e Zé Neto e Cristiano (que ficaram em terceiro lugar). Ao todo foram 15 batalhas ( $2^k - 1$ ), mas em apenas 5 delas houve diferença significativa entre os competidores: Marília Mendonça (2 vitórias), Jorge e Matheus (2 vitórias) e Israel e Rodolfo. A metodologia proposta permite inferir sobre a diferença existente entre competidores, e dessa forma, quais venceram por puro acaso.

É da natureza de todo trabalho científico ter limitações, ou seja, situações e condições não estudadas dentro de um tema. Este deixa uma série de sugestões de trabalhos futuros e pontos que devem ser aprofundados.

É possível formalizar um teste Bootstrap para a hipótese nula  $H_0 : a - b = 0$ . Essa hipótese podeira ser interpretada como a igualdade entre as forças dos competidores. Esse teste deve ser estudado quanto ao erro tipo I e poder.

Outro ponto é que os intervalos encontrados aqui são muito largos, devido às



observações terem vindo de uma única realização da batalha. Se as batalhas acontecerem simultaneamente em várias contas do Instagram, haverá uma amostra de proporções, o que pode reduzir a variância do estimador.

Para modelar as mesmas situações, outros estimadores podem ser comparados em desempenho, como o estimador de máxima verossimilhança, pelo método dos momentos (para  $n > 1$ ) e outros, como os estimadores robustos.

Por fim, o método dos momentos modificado deve ser formalizado como uma outra opção para obtenção de estimadores, e esses, por sua vez, devem ser estudados em termos suas propriedades (consistência, tendenciosidade, variância, erro quadrático médio, completude, etc.)

## REFERÊNCIAS

- ABIQ. **Associação Brasileira das Indústrias de Queijo**. 2022. [https://www.abiq.com.br/home\\_queijos.asp](https://www.abiq.com.br/home_queijos.asp). Acesso em: 28 ago. 2022.
- ADDO-PREKO, E.; AMISSAH, J. G. N.; ADJEI, M. Y. B. The relevance of the number of categories in the hedonic scale to the ghanaiian consumer in acceptance testing. **Frontiers in Food Science and Technology**, London, v. 3, 2023. ISSN 2674-1121. <https://www.frontiersin.org/articles/10.3389/frfst.2023.1071216>. Acesso em: 28 ago. 2022.
- AQUINO, A. A. F. **Atividades de modelagem matemática envolvendo a teoria dos grafos no ensino médio**. Dissertação (Mestrado em Matemática) — Universidade Estadual de Maringá, Maringá, PR, 2023.
- ARBELLO, D. D. R. **Parâmetros de qualidade do queijo colonial: percepção do consumidor & relação de consumo**. Dissertação (Mestrado em Ciência e Tecnologia dos Alimentos) — Universidade Federal de Santa Maria, Santa Maria, RS, 2021. Disponível em: <https://bit.ly/3knI2uC>. Acesso em: 28 Ago. 2022.
- BARBA, E. 2022. [@eric.barba]. (n.d.). Posts [Instagram profile]. Instagram. Disponível em: <https://www.instagram.com/eric.barba>. Acesso em: 16 ago. 2022.
- BASTOS, R. **Determinação de caminhos mínimos em aplicações de transporte público: um estudo de caso para a cidade de Porto Alegre**. Dissertação (Dissertação em Computação Aplicada) — Universidade do Vale do Rio dos Sinos, São Leopoldo, RS, 2013.
- BASTOS, R. L.; FERREIRA, E. B.; OLIVEIRA, M. S. d.; FERREIRA, D. F.; SIMÃO, I. Application of a parametric bootstrap test for the sensory quality index of fresh-cut eggplant. **Sigmae**, v. 3, n. 2, p. 34–38, Jan. 2015. Disponível em: <https://publicacoes.unifal-mg.edu.br/revistas/index.php/sigmae/article/view/341>. Acesso em 16 ago. 2023.
- BRASILIENSE, D.; SEIXAS, L. Sofrência em tempos de felicidade: música sertaneja e os signos da contemporaneidade. **Comunicação & Inovação**, São Caetano do Sul, SP, v. 21, n. 45, 2020.
- BURDEN, R. L.; FAIRES, J. D. **Numerical Analysis**. 9. ed. London: Brooks/Cole, Cengage Learning, 2010.
- CARNEIRO, J. P.; VILLAS BOAS, M. Cheetos brasil no instagram: O uso do seu mascote como porta-voz da marca. **Caderno de Estudos em Publicidade e Jornalismo**, Juiz de Fora, MG, v. 1, n. 1, 2019. Disponível em: <https://shorturl.at/cvDY9>. Acesso em: 28 ago. 2022.
- CARRASCO, J.; FABIO, L. C.; SANTANA, L. Métodos bootstrap: uma aplicação em dados olímpicos via python. **Revista Brasileira de Iniciação Científica**, Itapetininga, SP, p. e022014–e022014, 2022.
- CASELLA, G.; BERGER, R. L. **Statistical Inference**. Pacific Grove, CA: Duxbury Press, 2002.

CERRONI, S.; NOTARO, S.; RAFFAELLI, R. Beliefs and preferences for food-safety policies: a discrete choice model under uncertainty. **European Review of Agricultural Economics**, Oxford University Press, v. 46, n. 5, p. 769–799, 2019.

DAVISON, A. C.; HINKLEY, D. V. **Bootstrap methods and their application**. Cambridge: Cambridge university press, 1997.

DENSKI, A. P. N.; BACK, Á. J. Aderência de distribuições de probabilidade aos dados de evapotranspiração de referência decendial. **Revista Brasileira de Climatologia**, Curitiba, v. 17, fev. 2021. Disponível em: <https://ojs.ufgd.edu.br/index.php/rbclima/article/view/13842>. Acesso em: 10 ago. 2022.

EFRON, B. The 1977 rietz lecture. **The annals of Statistics**, Waite Hill, Ohio, v. 7, n. 1, p. 1–26, 1979. Disponível em: [http://xyala2.bio.ed.ac.uk/teaching/tutorials/phylogenetics/Bayesian\\_Workshop/PDFs/Effron%20Ann%20Stat%201979.pdf](http://xyala2.bio.ed.ac.uk/teaching/tutorials/phylogenetics/Bayesian_Workshop/PDFs/Effron%20Ann%20Stat%201979.pdf). Acesso em: 10 jul. 2022.

FARIA, A. H. P. de; ALVES, D. F. C.; BARROSO, L. C. Aplicação da Teoria de Grafos e análise espacial para solução de problemas geográficos: um estudo da criminalidade violenta no hipercentro de Belo Horizonte. *In: XVI Encontro da Associação Nacional de Pós-Graduação e Pesquisa em Planejamento Urbano e Regional*. **Anais [...]**, v. 16, n. 1, 2015.

FERREIRA, E. B.; VALADARES, F. L.; SILVA, C. P. d. Battle of cheeses: Tournament via instagram stories. **Sigmae**, Alfenas, v. 12, n. 1, p. 108–115, Feb. 2023. Disponível em: <https://publicacoes.unifal-mg.edu.br/revistas/index.php/sigmae/article/view/cheese>. Acesso em: 01 jul. 2023.

FERREIRA, M. C.; GONÇALVES, J. P. Educação e cultura popular: Representações de gênero na música sertaneja. **Humanidades & Inovação**, Palmas, TO, v. 8, n. 58, p. 341–359, 2021.

GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian data analysis**. London: Chapman and Hall/CRC, 1995. 668 p. ISBN 1-58488-388-X.

GODOY, M. M.; RODRIGUES NETO, J. Análise sensorial de queijo de leite de ovelha Santa Inês. **Veterinária e Zootecnia**, Botucatu, v. 29, dez. 2022. Disponível em: <https://rvz.emnuvens.com.br/rvz/article/view/1018>. Acesso em: 03 dez. 2022.

GUPTA, A. A. A new approach to bracket prediction in the NCAA men's basketball tournament based on a dual-proportion likelihood. **Journal of Quantitative Analysis in Sports**, Germany, v. 11, n. 1, p. 53–67, 2015.

GUPTA, A. K.; NADARAJAH, S. **Handbook of beta distribution and its applications**. London: CRC press, 2004.

GUYON, J. Rethinking the fifa world cup™ final draw. **Journal of Quantitative Analysis in Sports**, De Gruyter, v. 11, n. 3, p. 169–182, 2015.

HAMILTON, J.; MCILVEEN, H.; STRUGNELL, C. Educating young consumers—a food choice model. **Journal of Consumer Studies & Home Economics**, Wiley Online Library, v. 24, n. 2, p. 113–123, 2000.

LEEMIS, L. M. Relationships among common univariate distributions. **The American Statistician**, Taylor & Francis, v. 40, n. 2, p. 143–146, 1986.

LEITNER, C.; ZEILEIS, A.; HORNIK, K. Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the euro 2008. **International Journal of Forecasting**, London, v. 26, p. 471–481, 2010.

LEMOS, A.; SENA, C. D. Mais livre para publicar: efemeridade da imagem nos modos “galeria” e “stories” do instagram. **Revista Mídia e Cotidiano**, Rio de Janeiro, v. 12, n. 2, p. 6–26, 2018.

LIM, J. Hedonic scaling: A review of methods and theory. **Food Quality and Preference**, United States, v. 22, n. 8, p. 733–747, 2011. ISSN 0950-3293.

AGROSTAT 2010. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0950329311000954>. Acesso em: 16 set. 2022.

MACEDO, C. A. A.; ALBUQUERQUE, A. A. d.; MORALLES, H. F. Análise de viabilidade econômico-financeira de um projeto eólico com simulação monte carlo e avaliação de risco. **Gestão & Produção**, Universidade Federal de São Carlos, v. 24, n. 4, p. 731–744, Oct 2017. ISSN 0104-530X. Disponível em: <https://doi.org/10.1590/0104-530X3439-16>. Acesso em: 16 set. 2022.

MAQUENGO, G. L. **Teoria dos grafos e aplicações: redes eléctricas e de transportes rodoviários**. Dissertação (Mestrado em Matemática e Aplicações), Universidade de Évora, Portugal, 2019.

MOOD, A.; GRAYBILL, F.; BOES, D. **Introduction to statistical theory**. New York: McGraw-Hill, 1974.

PERYAM, D. R. Problem of preference gets GM focus. **Food Industries**, London, v. 22, n. 12, p. 42–44, 1950.

PERYAM, D. R.; PILGRIM, F. J. Hedonic scale method of measuring food preferences. **Food technology**, Chicago, n. 11, p. 9–14, 1957.

RAMOS, P. E. G. T.; MARTINS, A. de O. Reflexões sobre a rede social instagram: do aplicativo à textualidade. **Texto Digital**, Florianópolis, v. 14, n. 2, p. 117–133, 2018.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2023.

RESENDE, J. L. M. **Análise das distribuições de times campeões em um torneio do tipo mata-mata segundo a força dos times**. 2018. Monografia (Graduação em Ciências Econômicas) Universidade Federal de Ouro Preto, Ouro Preto, MG, Brazil. <https://www.monografias.ufop.br/handle/35400000/1110>.

RIBATEJO, A. **Tendências do Mercado Alimentar do Brasil: Aspectos do consumo, produto, distribuição e comunicação**. 2015. Disponível em: <https://bit.ly/3khHJBE>. Acesso em: 28 Ago. 2022.

SA, S. P. de. Cultura digital, videoclipes e a consolidação da rede de música brasileira pop periférica. **Fronteiras - estudos midiáticos**, São Leopoldo, RS, v. 21, n. 2, 2019.

SAMG. **Raio X: no Dia Mundial do Queijo conheça os números da produção em Minas**. 2022. Secretaria de Agricultura do Estado de Minas Gerais. Disponível em: <http://bit.ly/3ZjhERz>. Acesso em: 22 Jan. 2022.

SANTOS, A. B.; OZEKI, F. L.; OLIVEIRA, B. M. G.; KIMURA, M. Precisão de escalas de mensuração utilizadas em testes de aceitação. **Alimentos e Nutrição**, Araraquara, v. 20, n. 4, p. 633–640, 2010.

SILVA, C. M. da; VIEIRA, J. W.; NETO, N. A.; GONÇALVES, G. L. Controle estatístico de qualidade usando o método bootstrap. **Scientia Plena**, São Cristóvão, SE, v. 14, n. 3, 2018.

SILVA, F. A. da. Análise histórico-cultural da música sertaneja no brasil: do caipira ao playboy. **Revista de literatura, história e memória**, Cascavel, v. 14, n. 24, p. 218–234, 2018.

SILVA, R. D. G. da. Um olhar feminino na música sertaneja: aspectos do discurso e dos valores do feminejo. **Brazilian Journal of Development**, Curitiba, v. 7, n. 2, p. 18616–18628, 2021.

SINGRODIA, V.; MITRA, A.; PAUL, S. A review on web scrapping and its applications. In: **International Conference on Computer Communication and Informatics (ICCCI)**. Coimbatore, India: IEEE, 2019. p. 1–6.

SIQUEIRA, K. B. **O mercado de leite e derivados**. Juiz de Fora, MG: EMBRAPA, 2019. Circular Técnica 120. ISSN: 1678-037X. Disponível em: <https://bit.ly/3XU2EIw>. Acesso em 28 ago. 2022.

\_\_\_\_\_. Um retrato do consumo de lácteos no brasil. **Indústria de Laticínios**, n. 150, p. 58–59, 2021. Disponível em: <https://shorturl.at/fwA46>. Acesso em: 28 Ago. 2022.

SOARES, D. de A.; SOUZA, S. de A.; SILVA, D. José da; SILVA, A. B.; CAVALCANTE, U. M. B.; LIMA, C. M. B. L. Avaliação epidemiológica da esquistossomose no estado de pernambuco através de um modelo de regressão beta. **Archives of Health Sciences**, São José do Rio Preto, v. 26, n. 2, p. 116–120, nov. 2019. Disponível em: <https://ahs.famerp.br/index.php/ahs/article/view/110>. Acesso em: 12 ago. 2022.

SOARES, N. D.; CAMPOS, E.; NOGUEIRA, T.; SIQUEIRA, K.; DAVID, J. M. N.; BRAGA, R. Quais os tipos de queijos mais comentados no twitter. In: **Anais do Workshop de Iniciação Científica da EMBRAPA Gado de Leite**. Juiz de Fora, MG: EMBRAPA, 2021. Disponível em: <https://shorturl.at/ouIUW>. Acesso em: 28 Ago. 2022.

SOUZA, S. d. A.; SILVA, A. B.; CAVALCANTE, U. M. B.; LIMA, C. M. B. L.; SOUZA, T. C. d. Obesidade adulta nas nações: uma análise via modelos de regressão beta. **Cadernos de Saúde Pública**, Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz, v. 34, n. 8, p. e00161417, 2018. ISSN 0102-311X. Disponível em: <https://doi.org/10.1590/0102-311X00161417>. Acesso em: 18 ago. 2022.

SPOTIFY. **Top 50 Brazil**: seu relatório diário das faixas mais tocadas no momento. 2023. Acesso em: 25 mar. 2023. Disponível em: <https://spotify.link/HBdWJh8P4Db>.

USDA. **Dairy Market Statistics. 2020 Annual Summary**. 2020. United States Department of Agriculture. Agricultural Marketing Service. Disponível em: <https://www.ams.usda.gov/sites/default/files/media/DY2021Annual%20Summary.pdf>. Acesso em 5 set. 2022.

VIEIRA, S. M.; BANDEIRA, S. F. Desenvolvimento e avaliação da aceitabilidade de petisco de queijo. **Brazilian Journal of Development**, Curitiba, v. 6, n. 4, p. 20187–20191, 2020.

WICHCHUKIT, S.; O'MAHONY, M. The 9-point hedonic scale and hedonic ranking in food science: some reappraisals and alternatives. **Journal of the Science of Food and Agriculture**, London, v. 95, n. 11, p. 2167–2178, 2015. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jsfa.6993>.