



MICHAEL JACKSON OU RIHANNA? CÁLCULO DE UM ÍNDICE DE POPULARIDADE NA INDÚSTRIA MUSICAL

Gabriel da Costa Almeida¹ e Daniel Takata Gomes²

Introdução

As tecnologias digitais vêm modificando muito o modo como bens culturais são consumidos, em especial produtos musicais (Molteni & Ordanini, 2003). Até os anos 90, o sucesso de um artista era mensurado basicamente em termos de venda de álbuns físicos. A partir de então, com o advento da compressão digital de músicas em formatos de arquivos de computador e com o avanço das conexões de internet, serviços como compras de álbuns em formatos digitais e acesso a músicas em formato de streaming se popularizaram. Consequentemente, as vendas de álbuns físicos caíram, não sendo mais o único parâmetro do sucesso de um artista (Sandulli, 2007).

De acordo com dados do site ChartMasters.com, o álbum mais bem sucedido de Lady Gaga, *The Fame*, vendeu cerca de 15 milhões de cópias, enquanto o mais bem sucedido de Madonna, *True Blue*, vendeu cerca de 22 milhões de cópias. Por esse parâmetro, o álbum da segunda é mais bem sucedido. No entanto, as faixas do álbum de Lady Gaga foram baixadas quase 80 milhões de vezes pela internet, enquanto as faixas do álbum de Madonna foram baixadas cerca de 3 milhões de vezes. Não faz sentido simplesmente somar os números para determinar qual é o álbum mais popular. O mundo mudou, assim como o modo de consumir música.

Outro problema é a acurácia dos números. Como detalha o site ChartMasters.com, mesmo grandes veículos como a Forbes divulgam números distorcidos, na maioria das vezes superestimados. O referido, desde 2015, tem feito um trabalho para desmistificar as vendas dos artistas mais populares, trazendo dados confiáveis e números mais próximos da realidade.

O mesmo site apresenta uma medida denominada Commensurate Sales to Popularity Concept (CSPC), com o objetivo de ser uma escala consistente e relevante que considere

¹ Escola Nacional de Ciências Estatísticas (ENCE), gcosta12almeida@gmail.com

² Universidade Estadual de Campinas (UNICAMP), daniel.gomes@ibge.gov.br



todos os formatos de venda de música de uma forma justa. Por exemplo, a escala considera que o acesso a 150 streamings de áudio se iguala à venda de 1 single, e 10 singles vendidos se igualam a 1 álbum. Tais relações e ponderações são feitas entre todos os tipos de produtos musicais. No entanto, tais ponderações podem ser arbitrárias. O ideal é que os dados falem por si só.

Objetivos

O objetivo deste projeto é utilizar métodos de análise multivariada, em especial análise componente principal, para determinar um índice de popularidade dos artistas. Eventualmente, pode-se através deste estudo determinar índices de popularidade de álbuns e até mesmo de canções.

Material e Método

Informações sobre vendas de álbuns físicos e digitais fornecidos pelo site Chart Masters.com são utilizados na análise e para a montagem da base de dados. Os levantamentos feitos pelo site tentam, na medida do possível, evitar distorções frequentemente encontradas nos números oficiais, que muitas vezes são superestimados.

Cinco variáveis são consideradas, relacionadas aos respectivos artistas (total de 104) e álbuns: vendas de álbuns de estúdio, vendas de outros lançamentos, vendas de singles, streamings e downloads.

Análise de componentes principais tem como objetivo explicar a estrutura de variância-covariância de um conjunto de variáveis através de poucas combinações lineares dessas variáveis (Johnson & Wichern, 2007). Essas combinações lineares têm sido usadas em diversos trabalhos como uma forma de classificação, ou ranking, das unidades observacionais (Dawkins, B.1989; Kaycheng, S.2015).

O objetivo da aplicação da metodologia é buscar uma medida de pontuação que classificaria os artistas com base nos dados de vendas considerados.

Todas as análises foram executadas no software R (R Core Team,2018). Além disso, para a montagem da base foram utilizados os pacotes stringr (Wickham, 2019), readxl (Wickham & Bryan, 2018), dplyr (Wickham et al ,2019) e magrittr (Bache e Wickham, 2014).

Resultados e Discussão

Pelo fato de as variáveis apresentarem magnitudes diferentes, opta-se por padronizalas para efetuar a análise de componentes principais.



Nota-se que percentual da variância explicado pelas duas primeiras componentes principais em ambos os casos é de quase 73%. No caso de todos os coeficientes serem positivos, é a componente que pode ser utilizada como o índice de popularidade, e a componente com sinais alternados é a que separa vendas físicas de digitais.

Sendo assim, para a análise dos números dos artistas, observa-se que a componente mais indicada para efetuar a classificação desejada é a segunda componente principal, pois se trata de uma média ponderada das variáveis disponíveis. Logo, tomam-se tais valores e calcula-se a combinação linear referente a essa componente para cada artista, tendo, assim, uma medida de classificação. De forma análoga, foi feito o mesmo procedimento para o cálculo em relação aos álbuns dos artistas, dessa vez utilizando a primeira componente principal para a classificação. Os resultados se encontram na Tabela 1.

Por limitações de espaço, outras análises não são mostradas aqui, como análises gráficas e descritivas, distância de Mahalanobis e a interpretação das outras componentes principais.

Tabela 1 - Top 10 Artistas/Álbuns

Artista	Valor da 2ª comp. princ.	Álbum	Valor da 1ª comp. princ..
Rihanna	4.38	Thriller (Michael Jackson)	12.29
The Beatles	3.07	÷ (Ed Sheeran)	7.39
Michael Jackson	2.89	21 (Adele)	7.39
Ed Sheeran	2.61	The Fame (Lady Gaga)	7.28
Justin Bieber	2.55	Hotel California (Eagles)	6.77
Eminem	2.48	x (Ed Sheeran)	6.46
Elvis Presley	2.41	Bad (Michael Jackson)	6.23
Beyonce	2.25	Saturday Night Fever-(Bee Gees)	5.93
Taylor Swift	1.80	Rumours (Fleetwood Mac)	5.93
Queen	1.72	Purpose (Justin Bieber)	5.71

Conclusão

Com o objetivo de buscar um índice de popularidade na indústria musical, o presente trabalho propôs a aplicação da análise de componentes principais como forma de classificar os artistas com base nos números de vendas e downloads de seus álbuns e músicas.



Pode-se observar, em relação aos resultados, uma interessante mescla de artistas que se destacaram nas últimas décadas com outros que tiveram seus auge no século passado.

Referências

Dawkins, B. (1989). Multivariate analysis of national track records. *The American Statistician*, 43, 110-115.

Hadley Wickham (2019). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.4.0. <https://CRAN.R-project.org/package=stringr>

Hadley Wickham & Jennifer Bryan (2018). *readxl: Read Excel Files*. R package version 1.1.0. <https://CRAN.R-project.org/package=readxl>

Hadley Wickham, Romain François, Lionel Henry & Kirill Müller (2019). *dplyr: A Grammar of Data Manipulation*. R package version 0.8.0.1. <https://CRAN.R-project.org/package=dplyr>

Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. 6th ed., Prentice Hall, New Jersey.

Kaycheng, S. (2015). Multicollinearity and indicator redundancy problem in world university rankings: an example using Times Higher Education world university ranking 2013-2014 data. *Higher Education Quarterly*, 69 (2), 158-174.

Molteni, L., & Ordanini, A. (2003). Consumption patterns, digital technology and music downloading. *Long Range Planning*, 36 (4), 389-406.

R Core Team (2018). *R: A language & environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Stefan Milton Bache and Hadley Wickham (2014). *magrittr: A Forward-Pipe Operator for R*. R package version 1.5. <https://CRAN.R-project.org/package=magrittr>

Sandulli, F. D. (2007). CD music purchase behaviour of P2P users. *Technovation*, 27 (6-7), 325-334.