

Análise Estatística Multivariada de dados de XRF obtidos de Moedas Históricas utilizando Python

B. L. O. Gonçalves¹, M. L. Silva¹, M. B. Oliveira¹, M. A. Oliveira, R. P. Freitas¹

¹ barblorrane@gmail.com, IFRJ
¹ mariane.teceletro@gmail.com, IFRJ
¹ matheus.batista@ifrj.edu.br, IFRJ
¹ miguel.ao1086@gmail.com, IFRJ
¹ renato.freitas@ifrj.edu.br, IFRJ

1. Introdução

Nos últimos anos, o exame científico de obras de arte e objetos arqueológicos – denominado arqueometria – adquiriu interesse crescente, promovendo a interação entre campos distintos da ciência em torno de um propósito comum. A abordagem multidisciplinar de tais investigações científicas proporciona dados sobre a composição química, a tecnologia de fabricação, o período histórico e o estado de conservação destes artefatos [1,2].

Nesse contexto, a não destrutividade é um requisito essencial para os métodos científicos que operam em pesquisa no campo do patrimônio cultural: as informações obtidas pela investigação são dirigidas não só para o conhecimento completo de uma obra de arte, mas também para a sua melhor conservação [3,4]. Respeitando esse critério de análises não destrutivas, a Fluorescência de Raios X (XRF) tem sido um método de análise frequentemente aplicado na arqueometria, na investigação de artefatos arqueológicos [5].

É comum que os dados de XRF obtidos desses artefatos sejam submetidos a métodos estatísticos multivariados, como a Análise de Componentes Principais (PCA), que permitem avaliar semelhanças existentes entre as amostras a partir de um gráfico plotado com os “scores” do PCA [6,7]. Contudo, uma das dificuldades inerentes a essas análises empregando dados de XRF e PCA, advém do fato de que nesse tipo de investigação são realizadas medidas de muitas amostras, e os dados dessas medidas necessitam passar por tratamentos, cujas etapas são efetuadas singularmente para cada espectro de XRF, antes da análise por PCA. Esse processamento de dados geralmente demanda bastante tempo e, na maioria das vezes, exige o uso de mais de um *software* comercial. Portanto, o presente trabalho teve como principal objetivo desenvolver rotinas computacionais baseadas na linguagem Python, que pudessem facilitar e automatizar o estudo estatístico multivariado por PCA de dados de XRF, registrados em análises de 113 moedas históricas pertencentes ao Museu Histórico Nacional (MHN). Além disso, essas análises também auxiliarão os numismatas do MHN a realizar uma catalogação mais sistemática dos artefatos.

2. Metodologia

Inicialmente, foram feitas 3 medidas por XRF em cada uma das 113 moedas históricas do período Romano, datadas entre 43 a.C e 33 a.C, pertencentes ao MHN. O equipamento de XRF utilizado foi o modelo Tracer, da Companhia Bruker. Seu tubo de raios x é composto por um anodo de ródio que pode operar com tensão e corrente máximas de 40 kV e 200 μ A, respectivamente. Para as amostras em questão, cada espectro foi coletado durante um tempo de 60 segundos, com o tubo operando à tensão de 40 kV e corrente de 10 μ A. Na Fig. 1 é mostrada a imagem de uma das moedas Romanas analisadas.



Figura 1: Foto das faces de uma das moedas romanas investigadas.

Os dados espectrais adquiridos são contagens relacionadas à intensidade dos elementos detectados nas

moedas, e foram armazenados em arquivos com extensão *txt*. As rotinas computacionais desenvolvidas permitiram realizar os seguintes processamentos a partir desses arquivos de dados: cálculo da média das contagens de cada canal a partir dos espectros coletados em cada amostra; normalização das médias entre 0 e 1; construção da matriz de dados (2048 x 113) para entrada dos cálculos por PCA; desenvolvimento do método por PCA; construção de gráficos (*scores* e *loadings*) para visualizar os resultados produzidos por PCA.

PCA caracteriza-se como uma técnica de estatística multivariada que consiste em transformar um conjunto grande de dados em componentes principais (PC1, PC2, PC3...), que carregam o máximo de informação possível do conjunto original, reduzindo a massa de dados do estudo em questão.

Para a construção do código em Python, utilizou-se principalmente de quatro biblioteca: OS [8], Numpy [9], Scikit-learn [10] e Matplotlib [11]. O uso dessas bibliotecas justifica, também, a escolha dessa linguagem de programação, que demonstrou ser relativamente simples, completa e eficiente no que diz respeito a aplicações científicas, como no presente trabalho.

3. Resultados e Discussão

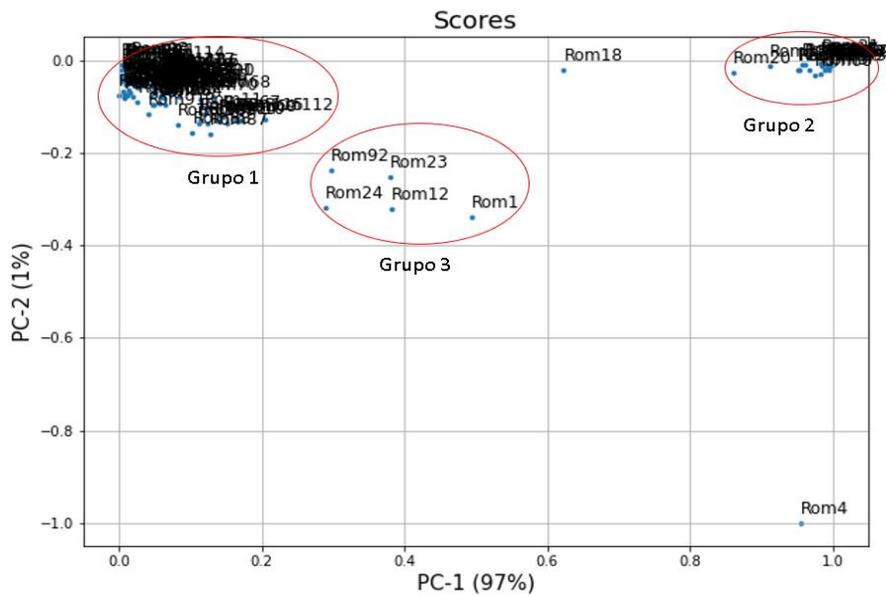


Figura 2: Gráfico dos *scores* do PCA das 113 moedas históricas.

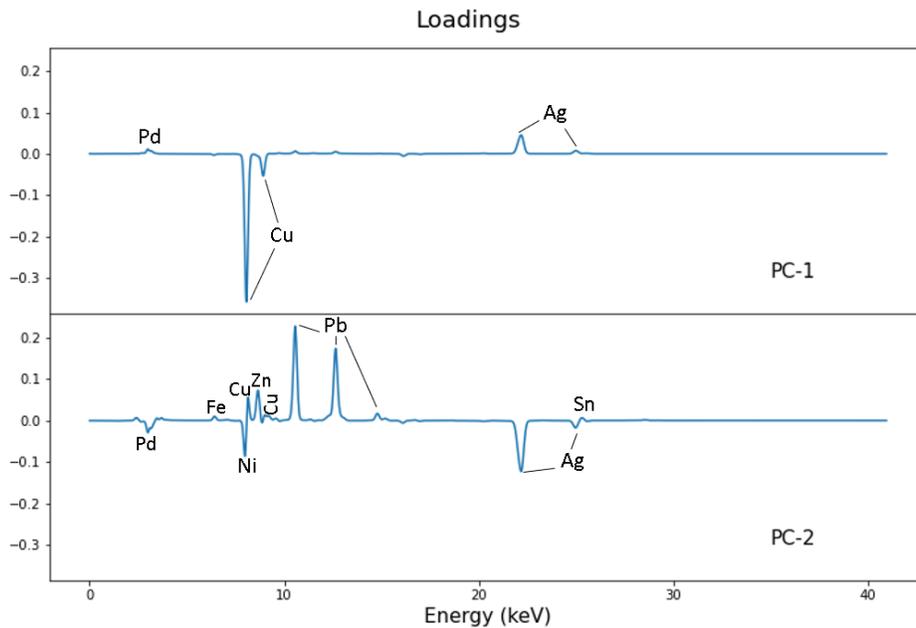


Figura 3: Gráfico de *loadings* referente às componentes 1 e 2 (PC-1 e PC-2).

O código, desenvolvido com o objetivo de automatizar o estudo estatístico multivariado por PCA, produziu o gráfico de *scores* (Fig. 2) como resultado central, e o de *loadings* (Fig. 3) como complementar à análise. No que se refere à interpretação dos *scores*, é possível visualizar 2 principais conjuntos de amostras, os quais estão identificados na Fig. 2 como Grupo 1 e Grupo 2. Além disso, identificou-se ainda um terceiro grupo, contendo neste somente 5 amostras, e duas amostras *outliers* (Rom 4 e Rom 18). Como comentado na seção anterior, os agrupamentos das amostras indicam similaridades entre elas, nesse caso, em termos da composição elementar. Nos *scores*, os dois primeiros componentes principais (PC-1 e PC-2) correspondem respectivamente a 97% e 1% da variância observada, totalizando 98%. Essa variância ocorre pelas diferenças na composição elementar base da liga metálica das moedas, possivelmente por conta do processo de manufatura ou na utilização de matéria-prima distinta.

O gráfico de *loadings plots* (Fig. 3) explicitam como as intensidades dos elementos químicos identificados interferem na distribuição das amostras no gráfico dos *scores* do PCA. Esses gráficos possibilitam realizar um estudo mais profundo do resultado do PCA, sendo esse tipo de investigação fundamental no processo de compreensão das diferenças apresentadas pelas amostras e identificação das razões pelas quais elas se agrupam e/ou se separam. As análises realizadas por meio desses gráficos (*scores* e *loadings*), combinadas às informações obtidas através de alguns espectros de XRF dessas moedas (Fig. 4), permitem a inferência de que Cu, Zn e Fe, nessa ordem, são os elementos de maior intensidade nas amostras pertencentes ao Grupo 1. Já para o Grupo 2, as análises apontaram os elementos Ag, Cu e Pd como os que mais influenciam este grupo. Para as amostras do Grupo 3, Cu, Pb e Fe são os elementos de maior intensidade. Por fim, os destacamentos encontrados nas amostras *outliers* ocorrem por conta da amostra Rom 4 apresentar uma grande concentração de Pb em sua composição, maior do que em qualquer outro grupo de amostras, bem como a amostra Rom 18 apresentar uma razoável concentração de Cu, predominante no Grupo 1, entretanto apresentar também considerável concentração de Ag, elemento muito presente nas amostras do Grupo 2, caracterizando essa moeda como uma amostra “meio-termo” entre esses dois grupos, mas que não se enquadra completamente em nenhum deles.

Na Fig. 4 são mostrados alguns espectros de XRF referentes a amostras pertencentes a cada um dos grupos, bem como das amostras *outliers*. A visualização desses espectros permite perceber como cada grupo definido é caracterizado, e contribuiu para as análises junto aos gráficos anteriores (Fig. 2 e 3), a partir das quais foi possível observar que, de maneira geral, os principais elementos identificados nas moedas são Ag, Cu, Fe, Pb, Pd e Zn. Uma observação pertinente é a de que, nas análises espectrais, verifica-se que as linhas da camada L do Pd aparecem sobrepostas às linhas L do Rh, sendo esse segundo detectado por compor o anodo do tubo de Raios X do equipamento, e por isso não é indicado nas análises.

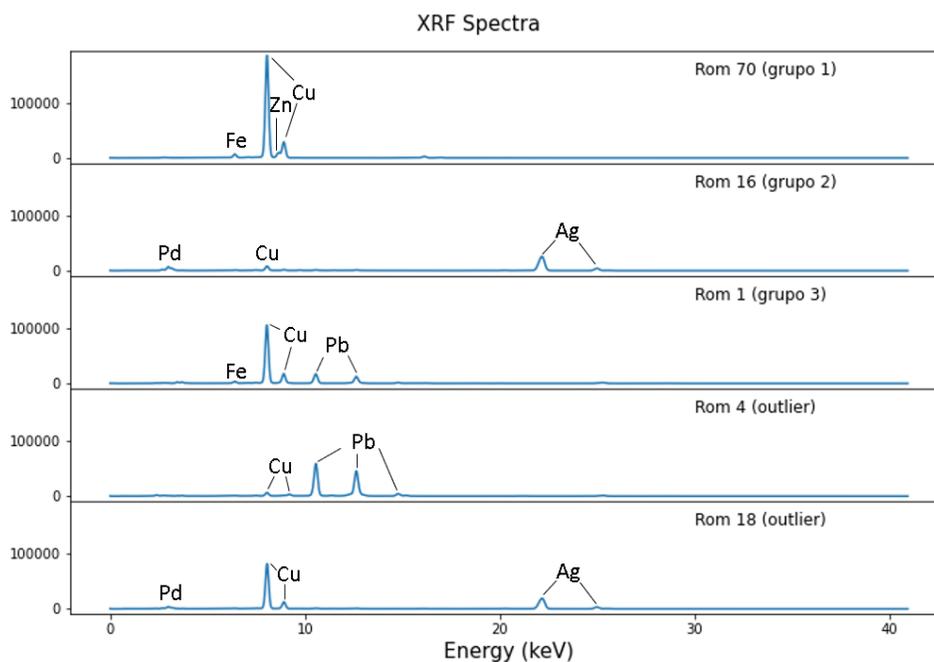


Figura 4: Espectro XRF das amostras

4. Conclusão

Para verificar o funcionamento do código construído, os resultados obtidos foram comparados aos de um *software* comercial, o The UnscramberX. Para produzir o gráfico dos *scores* do PCA por meio do código, foi necessário apenas aplicar o processamento de dados criado diretamente na pasta que contém os arquivos desejados, e dessa forma, todos os dados sofreram um pré-tratamento antes da plotagem do gráfico, tudo de forma automática. Contudo, para produzir o gráfico dos *scores* do PCA utilizando o software, foi necessário, anteriormente, acessar cada arquivo *txt* armazenado na pasta por vez, para tratá-los individualmente, e só depois submetê-los ao software para a construção do gráfico. Os resultados obtidos foram os mesmos. Portanto, pode-se afirmar que o código desenvolvido neste trabalho foi capaz de realizar todas as etapas necessárias para automatizar o estudo estatístico multivariado por PCA de dados de XRF, produzindo seu resultado central (gráfico dos *scores*) de maneira correta e rápida. Essa análise permite a obtenção de informações, como a composição elementar das amostras, que contribuem para a descoberta de similaridades entre os artefatos estudados e para a avaliação do processo de manufatura empregado. Portanto, todos esses aspectos puderam ser analisados após uma rotina de tratamento de dados completamente automatizada, e o presente estudo cumpriu o seu papel de otimizar o processo em questão.

Referências

- [1] T.A. Lima, C. Calza, S.M. de Souza, A. Brancaglione Jr, M.J. Anjos, M.I.M.S. Bueno, R.T. Lopes, XRF applications in archaeometry: analysis of Marajoara pubic covers and pigments from the sarcophagus cartonnage of an Egyptian mummy, *X-Ray Spectrom.*, vol. 36, pp. 348-354 (2007). doi:10.1002/xrs.982.
- [2] R.P. Freitas, Aplicações de Técnicas Nucleares e Espectroscopia Molecular em Arqueometria, Programa de Engenharia Nuclear, COPPE, Universidade Federal do Rio de Janeiro, 2014.
- [3] D. Barilaro, G. Barone, V. Crupi, D. Majolino, P. Mazzoleni, G. Tigano, V. Venuti, FT-IR absorbance spectroscopy to study Sicilian “proto-majolica” pottery, *Vib. Spectrosc.*, vol. 48, pp. 269-275 (2008). doi:10.1016/j.vibspec.2008.01.005.
- [4] R.P. Freitas, I.M. Ribeiro, C. Calza, A.L. Oliveira, V.S. Felix, D.S. Ferreira, A.R. Pimenta, R.V. Pereira, M.O. Pereira, R.T. Lopes, Analysis of a Brazilian baroque sculpture using Raman spectroscopy and FT-IR, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 154, pp. 67-71 (2016). doi:10.1016/j.saa.2015.10.013.
- [5] L. Musilek, T. Čechák, T. Trojek, X-ray fluorescence in investigations of cultural relics and archaeological finds, *App. Radiat. Isot.*, vol. 70, pp. 1193-1202 (2012). doi:10.1016/J.APRADISO.2011.10.014.
- [6] R.P. Freitas, C. Calza, T.A. Lima, A. Rabello, R.T. Lopes, EDXRF and multivariate statistical analysis of fragments from Marajoara ceramics, *X-Ray Spectrom.*, vol. 39, pp. 307-310 (2010). doi:10.1002/xrs.1200.
- [7] R.P. Freitas, F.A. Coelho, V.S. Felix, M.O. Pereira, M.A.T. de Souza, M.J. Anjos, Analysis of 19th century ceramic fragments excavated from Pirenópolis (Goiás, Brazil) using FT-IR, Raman, XRF and SEM, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 193, pp. 432-439 (2018). doi:10.1016/j.saa.2017.12.047.
- [8] OS, <https://docs.python.org/3/library/os.html> (2021).
- [9] NumPy, <http://www.numpy.org/> (2021).
- [10] Scikit-learn, <https://scikit-learn.org/stable/> (2021).
- [11] Matplotlib, <https://matplotlib.org/#> (2021).