

UMA PROPOSTA DIDÁTICO-PEDAGÓGICA UTILIZANDO UM SOFTWARE LIVRE PARA INTERPRETAR A PCA

Rebeca Fernanda de Souza e Silva^a, Natália Neves Macedo Deimling^a e Patrícia Valderrama^{a,*} ^aDepartamento Acadêmico de Química, Universidade Tecnológica Federal do Paraná (UTFPR), 87301-899 Campo Mourão – PR, Brasil

Recebido: 12/12/2023; aceito: 19/04/2024; publicado online: 26/06/2024

A PEDAGOGIC-DIDACTICAL PROPOSAL BY USING FREE SOFTWARE TO INTERPRET PCA. Chemometrics is gaining more and more space in universities and industry. To contribute to the development and expansion of principal component analysis (PCA) teaching proposals, this work evaluated the answers of the students to a questionnaire and presented a didactic-pedagogical proposal that can contribute to solving the principal doubts achieved by students in the PCA interpretation. Based on data collected in a semi-open questionnaire, the proposal was developed based on methodological principles of historical-critical pedagogy, containing the objectives, the problems surrounding the topic, and the cultural instruments necessary for understanding and interpreting PCA in the discipline of chemometrics by using new resources and instruments, especially free software, Chemostat. This proposal covers topics related to the concept of chemometrics and PCA, the explanation of the data set, pre-processing, use of Chemostat software, and analysis of principal components. It is hoped that this proposal can promote a better understanding concerning PCA on the part of students, as well as effective incorporation of the phenomena that involve this tool and its applications, and interpretation. Likewise, it is expected that this didactic-pedagogical proposal can support new teaching practices in chemometrics subjects in higher education.

Keywords: chemometrics; PCA; higher education; pedagogical proposal.

INTRODUÇÃO

A quimiometria enquanto uma área de destaque da química analítica vem se mostrando de grande importância na interpretação e análise de dados químicos. Por definição, a quimiometria é a disciplina química que usa métodos matemáticos e estatísticos para planejar ou otimizar procedimentos experimentais e extrair o máximo da informação química relevante, através da análise destes dados.¹⁻³

Dentro da quimiometria existem várias subdivisões que têm por objetivo o planejamento de experimentos, a calibração multivariada e o reconhecimento de padrões.⁴ No reconhecimento de padrões destaca-se a ferramenta não supervisionada (ou exploratória) denominada análise de componentes principais (*principal component analysis* - PCA), que acaba se tornando base para muitas outras ferramentas da quimiometria. A PCA tem por objetivo visualizar estrutura de dados, encontrar similaridades entre as amostras, reduzir a dimensionalidade do conjunto de dados e detectar amostras anômalas (*outliers*).⁵

Considerando a importância que a quimiometria vem apresentando e o seu caráter interdisciplinar, sua disseminação acaba chamando bastante atenção dos alunos de graduação e pós-graduação que, muitas vezes, não dispõem de um conhecimento prévio sobre o tema, ou também sobre a utilização de *softwares*, indispensáveis para a sua aplicação. Dessa forma, o conteúdo é muitas vezes considerado bastante complexo por parte dos estudantes.

Com a popularização da quimiometria, considera-se importante a discussão sobre estratégias didático-pedagógicas que possam ser utilizadas para facilitar a compreensão dos estudantes sobre os conteúdos e ferramentas utilizadas nessa disciplina. Um panorama sobre o ensino de quimiometria nos cursos de graduação em química no Brasil foi traçado no trabalho de Guizzelli e Rossi.⁶ De acordo com as autoras, desde a publicação das Diretrizes Curriculares para os Cursos Superiores em Química, em 2001, foi possível a flexibilização na organização dos currículos por iniciativa das instituições de

ensino superior e publicações oficiais recomendando a abordagem de quimiometria nesses cursos. Apesar disso, a inserção formal desse conteúdo na matriz curricular de cursos de graduação em Química ainda é discreta se comparada com as tendências de pesquisas científicas crescentes sobre essa área e com a intensificação de suas aplicações industriais.⁷ Dessa forma, uma reflexão é merecida, já que se relaciona com a formação dos novos químicos para o mundo do trabalho.

Vale salientar que, no Brasil, existem trabalhos disponíveis na literatura⁸⁻¹⁸ propondo tutoriais ou materiais didáticos para o ensino de quimiometria. Dentre eles, destacam-se alguns que abordam a ferramenta PCA, como no trabalho de Lyra *et al.*⁸ Nele, o ensino da PCA é proposto a partir da análise das propriedades dos elementos da tabela periódica, enquanto que no trabalho de Valderrama *et al.*,¹⁸ o ensino da PCA é sugerido a partir da espectroscopia na região do ultravioleta e visível (UV-Vis) e dos equilíbrios químicos envolvidos nas transformações das antocianinas do repolho roxo em função do pH do sistema. Além desses, Souza e Poppi,⁹ indicam uma alternativa para ensinar PCA através de espectros na região do infravermelho médio de óleos vegetais. Outra alternativa é apresentada por Lopes *et al.*,¹⁹ que sugerem a identificação de marcas de canetas esferográficas. Na maioria dos casos, o emprego de *softwares* comerciais acaba por dificultar o desenvolvimento dessas propostas de ensino de quimiometria. Além disso, ao descreverem e discutirem a proposta de ensino, esses autores não explicitam a abordagem pedagógica utilizada.

Tendo em vista contribuir com o desenvolvimento e a aplicação de propostas didático-pedagógicas para o ensino de PCA, objetiva-se com este trabalho apresentar uma proposta de ensino de PCA nos cursos de ensino superior em que disciplinas de quimiometria estejam inseridas no currículo. Para essa proposta será considerado como referencial teórico-metodológico a pedagogia histórico-crítica.

A pedagogia histórico-crítica, elaborada por Dermeval Saviani,²⁰ é uma teoria da educação brasileira em construção desde a década de 1970. Ao resgatar a importância da educação escolar enquanto espaço de aquisição da cultura em sua forma mais elaborada, essa

*e-mail: patriciav@utfpr.edu.br; pativalderrama@gmail.com

teoria defende, entre outros aspectos, a importância de um processo educativo diretivo e intencional que favoreça a problematização e a instrumentalização dos conhecimentos científicos, os quais devem ser incorporados pelos estudantes para que superem uma visão puramente empírica de ciência e de mundo.

Trata-se de uma teoria que propõe, entre outros aspectos, um processo de ensino-aprendizagem que garanta os meios necessários para que os estudantes não apenas assimilem o saber objetivo enquanto resultado, mas que também apreendam o processo de sua produção. Tendo em vista superar uma visão fragmentada sobre a prática social, essa teoria também defende a importância de que o conhecimento seja abordado a partir das diferentes dimensões que compõem essa prática, entre elas aquelas de ordem conceitual, histórica, científica, política, entre outras necessárias para sua compreensão ampla, crítica e elaborada.²¹ Neste trabalho, foca-se em duas dimensões do conteúdo sobre PCA: a conceitual e a científica.

Considerando esses princípios, e tendo em vista o objetivo desse trabalho, este artigo apresenta uma proposta didático-pedagógica para o ensino de PCA em disciplinas de quimiometria no ensino superior, utilizando como ferramenta uma tabela de dados compiladas de um livro de bioquímica.

METODOLOGIA

Inicialmente, foi aplicado um questionário semiaberto aos estudantes que cursaram as disciplinas TAM58-quimiometria e ITC12-quimiometria, ministradas nos anos de 2019, 2020 e 2021 em dois programas de pós-graduação da Universidade Tecnológica Federal do Paraná (UTFPR): um em inovações tecnológicas e outro em tecnologia de alimentos. O questionário contemplava questões referentes à complexidade das aulas práticas sobre PCA, às dificuldades em relação à introdução dos dados, às contribuições de artigos como suporte para a explicação e interpretação dos dados, bem como às facilidades encontradas pelos estudantes nessas disciplinas e suas sugestões de aprimoramento do processo, em especial no que diz respeito às práticas de PCA. Ao todo, 26 alunos responderam ao questionário. Vale ressaltar que esses alunos realizaram aulas práticas sobre PCA no *software ChemoStat* e utilizando conjuntos de dados já publicados na literatura em outros tutoriais.^{8,18}

A partir desses dados, foi elaborada uma proposta didático-pedagógica baseada em alguns dos princípios metodológicos da pedagogia histórico-crítica contendo os objetivos, os problemas que envolvem o tema e os instrumentos culturais necessários para a compreensão e interpretação da PCA na disciplina de quimiometria a partir de novos recursos e instrumentos, em especial um *software* de uso livre, o *ChemoStat*.²²

Essa proposta contempla tópicos relacionados ao conceito de quimiometria e PCA e à explicação sobre conjunto de dados extraído do livro *Princípios de Bioquímica de Leningher*,²³ pré-processamento, utilização do *ChemoStat*.²²

A seguir, são apresentados os procedimentos para a interpretação dos dados da PCA a partir dessa nova proposta.

Conjunto de dados

O conjunto de dados utilizado nessa proposta consiste em valores tabelados de propriedades de aminoácidos naturais. Para a realização da PCA foram utilizados os valores tabelados das seguintes propriedades: massa resultante (M_r), valores de pK_1 e pK_2 que correspondem aos valores de pH onde o aminoácido funciona como um tampão durante uma curva de titulação; o ponto isoelétrico (pI) que é o pH em que um aminoácido tem uma carga líquida igual a zero (0), o índice hidropático (IH) que representa uma escala combinando hidrofobicidade e hidrofília e pode ser empregada para prever quais aminoácidos serão encontrados em um meio aquoso (valores negativos) ou um meio hidrofóbico (valores positivos),²⁴ e a ocorrência nas proteínas, calculada como uma média empregando-se valores de mais de 200 proteínas.²⁵ Foram empregados os aminoácidos com grupos R alifáticos e não polares, e os aminoácidos com grupos R aromáticos, conforme Tabela 1.

Os alunos são orientados a construir uma tabela de dados (matriz) em que os aminoácidos (objetos) são dispostos nas linhas da matriz, enquanto que as suas propriedades (variáveis) encontram-se nas colunas da referida matriz.

Pré-processamento

A matriz de dados construída a partir das propriedades dos aminoácidos possui variáveis com significados físicos, magnitudes e unidades distintas e, portanto, os dados precisam ser pré-processados. Neste caso o pré-processamento adequado é o autoescalamamento, que consiste em subtrair, de cada elemento da coluna da matriz, o valor médio da respectiva coluna e, em seguida, dividir o resultado pelo desvio padrão dessa coluna. Com esse pré-processamento todas as variáveis passam a exercer influências equitativas nos resultados.³

Programa computacional

Diferentes *softwares* estão disponíveis para realização de PCA. Neste trabalho, optou-se por utilizar o *software* livre *Chemostat*,²² que não necessita de instalação prévia. Entretanto, o professor poderá utilizar qualquer *software* que tenha disponível para ministrar esta

Tabela 1. Propriedades de aminoácidos naturais

Número da amostra/objeto	Aminoácido	M_r / kDa	pK_1	pK_2	pI	Índice hidropático	Ocorrência nas proteínas / %
1	glicina	75	2,34	9,60	5,97	-0,4	7,5
2	alanina	89	2,34	9,69	6,01	1,8	9,0
3	valina	117	2,32	9,62	5,97	4,2	6,9
4	leucina	131	2,36	9,60	5,98	3,8	7,5
5	isoleucina	131	2,36	9,68	6,02	4,5	4,6
6	prolina	115	1,99	10,96	6,48	-1,6	4,6
7	fenilalanina	165	1,83	9,13	5,48	2,8	3,5
8	tirosina	181	2,20	9,11	5,66	-1,3	3,5
9	triptofano	204	2,38	9,39	5,89	-0,9	1,1

Fonte: adaptado de Nelson e Cox.²³ M_r : massa resultante; pK_1 e pK_2 : correspondem aos valores de pH onde o aminoácido funciona como um tampão durante uma curva de titulação; pI: ponto isoelétrico.

aula. Certamente, os resultados e os gráficos que serão apresentados aqui poderão ser reproduzidos em outro *software*.

Análise de componentes principais

PCA é uma ferramenta quimiométrica de reconhecimento de padrões não supervisionado, também dito como exploratório. A PCA é capaz de transformar uma tabela de dados numéricos em gráficos informativos acerca da similaridade e diferenças entre as amostras e as respectivas variáveis responsáveis por isso.^{5,18} Essa tabela de dados numéricos, também denominada de matriz **X**, pode ser constituída a partir de dados discretos (valores de parâmetros físico-químicos, composições centesimais, etc.) ou contínuos (espectros na região do infravermelho, por exemplo). Embora alguns artigos dedicados a uma abordagem didática sobre a PCA estejam disponíveis na literatura,^{8,9,18} não foram encontradas abordagens didáticas da PCA aplicadas a dados discretos e utilizando o *software ChemoStat*.

A matriz de dados numéricos (matriz **X**) será decomposta na PCA em um produto de duas outras matrizes, uma matriz é denominada de *scores* ou *escores* (**T**) e outra matriz denominada de *loadings* ou pesos (**P**), mais uma matriz de erros (**E**). Nesta etapa ocorre uma redução de dimensionalidade representada por *d*, ou seja, o número de componentes principais (PCs) retidos na decomposição da matriz **X** por PCA, conforme a Equação 1:⁵

$$\mathbf{X} = \mathbf{T}_d \times \mathbf{P}_d^T + \mathbf{E} \quad (1)$$

Na decomposição da matriz **X** em *scores* e *loadings*, novos eixos denominados PCs, são calculados no espaço multidimensional dos dados numéricos. Os PCs explicam sucessivamente a maior percentagem de variância em uma dada direção e são ortogonais entre si. Essa redução na dimensionalidade dos dados numéricos agrupam as informações correlacionadas em uma mesma PC e, enquanto a matriz de *scores* (**T**) carrega informações acerca das semelhanças

ou diferenças entre amostras (linhas da matriz **X**), os *loadings* (**P**) fornecem informações sobre as variáveis (colunas de **X**), responsáveis por essas semelhanças ou diferenças encontradas nos *scores*. Além disso, os resultados das matrizes de *scores* e *loadings* podem ser interpretados graficamente.^{3,18}

RESULTADOS E DISCUSSÃO

Respostas do questionário

O Quadro 1 apresenta as perguntas, e as opções de respostas, utilizadas no questionário aplicado aos alunos.

A Figura 1 apresenta os resultados, em porcentagem, das respostas objetivas. Para o universo de alunos avaliados a etapa mais complexa para a aula prática sobre PCA está na interpretação dos resultados (Figura 1A).

Pensando na introdução dos dados no programa *ChemoStat*, 34,6% dos alunos consideraram que o número de conjuntos de dados utilizados nas aulas práticas (2 ou 3) foi satisfatório. Essa mesma porcentagem de alunos considerou a quantidade de conjunto de dados é indiferente para o processo de aprendizagem no que diz respeito a introdução dos dados no programa. Para o universo de alunos avaliados, 23,1% acreditam que empregar mais conjuntos de dados pode contribuir para o reforço da aprendizagem nesta etapa. Essa mesma porcentagem dos alunos considerou que conjuntos de dados extensos atrapalham o andamento da aula (Figura 1B).

Acerca do conjunto de dados utilizado no desenvolvimento da aula prática sobre PCA, 84,6% dos alunos responderam que consideram que ter suporte de artigos de apoio (tipo tutoriais) com explicações sobre os dados e os pré-processamentos pode facilitar o aprendizado (Figura 1C). Por fim, 84,6% dos alunos consideraram que trabalhar a interpretação dos resultados com base em artigos tutoriais promove uma aprendizagem mais satisfatória. Ainda, 76,9%

Quadro 1. Questionário aplicado aos estudantes da disciplina de quimiometria

Perguntas	Opções de resposta
1) Qual a etapa mais complexa do seu ponto de vista para a aula prática sobre PCA? Por favor, selecione apenas uma opção.	a) Introduzir os dados no programa <i>ChemoStat</i> . b) Executar os cálculos da PCA através do programa <i>ChemoStat</i> . c) Interpretar os resultados.
2) Com relação à introdução dos dados no programa <i>ChemoStat</i> você considera: Por favor, selecione todas as alternativas que achar pertinente.	a) O número de conjuntos de dados utilizados (2 ou 3) como satisfatório. b) Poderia usar mais conjuntos de dados. c) Os conjuntos de dados devem ser relativamente pequenos. d) Conjuntos de dados extensos atrapalham o andamento da aula. e) Indiferente.
3) Com relação aos conjuntos de dados utilizados nas aulas práticas sobre PCA: Por favor, selecione apenas uma das opções.	a) Considero que ter suporte de artigos de apoio (tipo tutoriais) com explicações sobre os dados e os pré-processamentos (Exemplo – na aula será empregado o conjunto de dados sugerido por esse artigo tutorial). b) Considero que ter suporte de artigos de apoio (tipo tutoriais) com explicações sobre os dados e os pré-processamentos é irrelevante. (Exemplo – na aula podemos utilizar qualquer conjunto de dados obtidos em artigos já publicados sem que sejam artigos tutoriais).
4) Com relação à interpretação dos resultados obtidos pela prática sobre PCA: Por favor, selecione todas as alternativas que achar pertinente.	a) Considero que trabalhar a interpretação dos resultados com base em artigos tutoriais promove uma aprendizagem mais satisfatória. b) As interpretações baseadas em artigos tutoriais são irrelevantes para a aprendizagem. c) As interpretações dos resultados da PCA são geralmente triviais e não são necessárias aulas práticas para aprender a realizar tais interpretações. d) O aprendizado acerca das interpretações dos resultados em aulas práticas sobre PCA é favorecido.
5) Relate as principais facilidades encontradas por você nas disciplinas de quimiometria, em especial no que se refere as aulas práticas sobre PCA.	Questão de resposta subjetiva.
6) A partir de sua experiência como estudante nessas disciplinas, apresente sugestões para o ensino de práticas de PCA em quimiometria	Questão de resposta subjetiva.

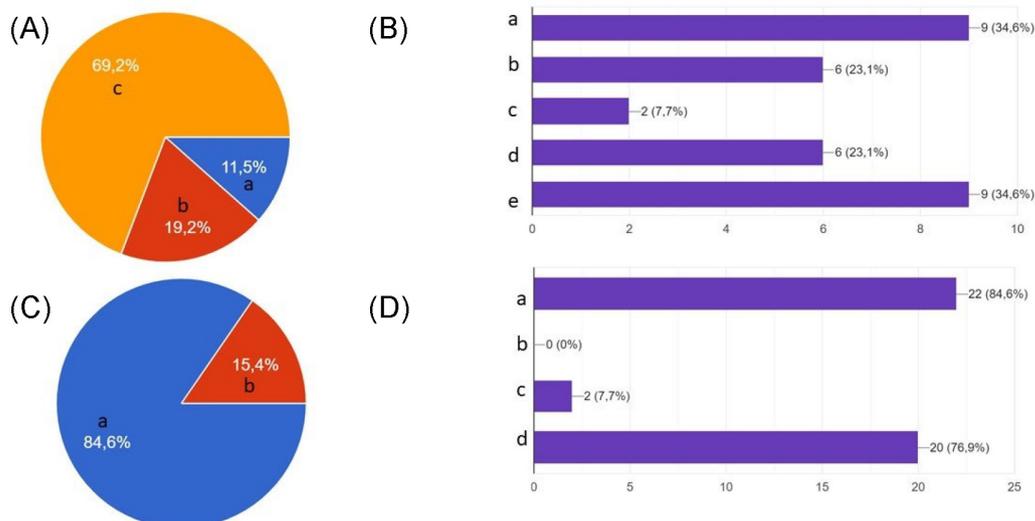


Figura 1. Respostas das perguntas objetivas do questionário. Respostas para a primeira (A), segunda (B), terceira (C) e quarta (D) perguntas

dos alunos relataram que o aprendizado acerca das interpretações dos resultados em aulas práticas sobre PCA é favorecido (Figura 1D).

Com relação às questões subjetivas, no que diz respeito às facilidades, os alunos consideraram o programa *ChemoStat* de fácil execução. Além disso, indicaram que as explicações e práticas propostas foram positivas para a aprendizagem do conteúdo teórico sobre PCA. Sobre as dificuldades encontradas, alguns estudantes relataram serem formados em outras áreas do conhecimento, o que acabou por prejudicar a compreensão e interpretação dos dados. Ao final, alguns estudantes indicaram algumas sugestões para uma melhor interpretação dos dados, tais como utilização de mais artigos como exemplos, mais exercícios e discussões sobre os resultados encontrados, ampliação do conjunto de dados para análise e aplicação de mais práticas com o uso do *software ChemoStat*, o que indica o potencial desse recurso de uso livre nas disciplinas de quimiometria.

Considerando esses dados, são apresentados a seguir os conhecimentos e as ações didático-pedagógicas necessárias para

a interpretação e discussão dos resultados da PCA na disciplina de quimiometria a partir dessa nova proposta, contemplando as dimensões conceitual e científica. O professor também pode optar por trabalhar as dimensões históricas e sócio ambientais se baseando no plano de unidade disponível no Material Suplementar.

PCA (*principal component analysis*)

O primeiro passo para a execução da PCA é a introdução dos dados no programa computacional. Nesta proposta sugere-se o *software* livre *ChemoStat*, disponível para download em www.chemostat.com.br.²² O usuário deve descompactar os arquivos e clicar no arquivo executável *ChemoStat.exe*, para abrir a janela apresentada na Figura 2a.

A partir da Figura 2a observa-se que o *software* permite a análise de dados contínuos como espectros, imagens digitais, mas também é possível a análise de dados discretos no ícone “import data”. Para importar os dados discretos, sugere-se que o usuário já possua a

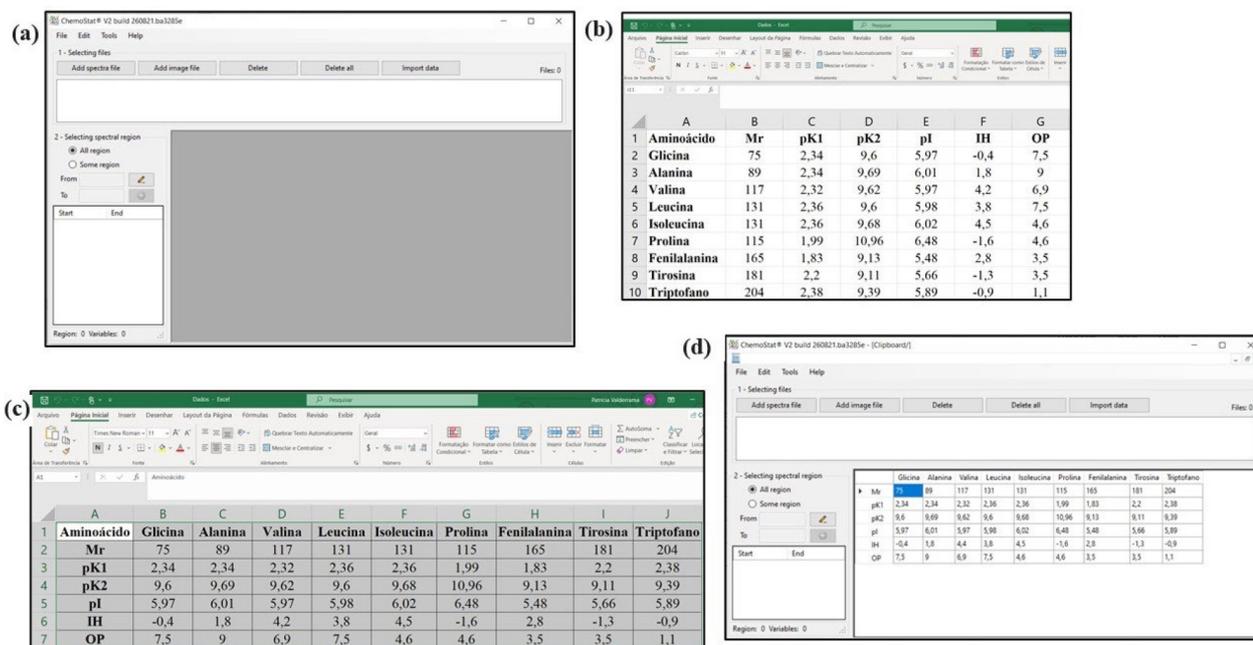


Figura 2. Iniciando o uso do *software ChemoStat*: (a) imagem inicial do *software*; (b) imagem do excel com os dados numéricos digitalizados no modo convencional; (c) imagem do excel com os dados numéricos digitalizados no modo correto para utilização no *ChemoStat*; (d) dados inseridos no *ChemoStat*

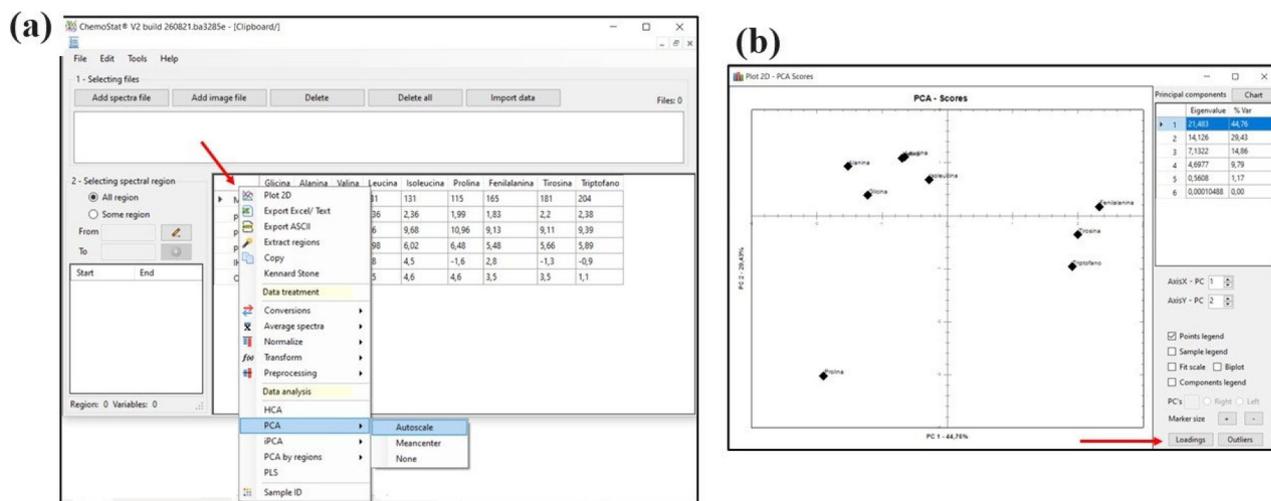


Figura 3. Usando o software ChemoStat: (a) como escolher a opção de análise PCA com o pré-processamento de autoescalamento e (b) gráfico dos scores de PC1 e PC2

tabela de dados numéricos pronta em outro arquivo, neste caso optou-se preparar a tabela de dados empregando o excel. É importante observar que a maioria dos softwares de análise quimiométrica opera utilizando o padrão de tabela da Figura 2b, com as amostras nas linhas e as propriedades (variáveis) nas colunas (formando uma matriz \mathbf{X}). Com o software ChemoStat utiliza-se as amostras nas colunas e as propriedades nas linhas, conforme Figura 2c (que seria a matriz \mathbf{X}^T). O usuário deve copiar toda a informação de interesse disponível na tabela (inclusive a parte de nome de amostras e variáveis) e após, na tela inicial do software ChemoStat, clicar no ícone “import data” para visualizar os dados importados para o software como mostra a Figura 2d. Neste ponto, vale ressaltar que é necessário configurar, previamente, a vírgula como ponto flutuante no arquivo excel. Para tanto, no arquivo excel clique em “arquivo”, selecione “opções”, na guia “avançado”, em “opções de edição” verifique a caixa de seleção “usar separadores do sistema” que deve estar com “,” (vírgula) no separador decimal e o “.” (ponto) no separador de milhar. Verifique também nas configurações do computador, em “hora e idioma” > “região” > “país ou região” deve estar Brasil e “formato de região” deve estar em [Português (Brasil)].

A Figura 3a mostra que, no espaço indicado com a seta vermelha, o usuário deve clicar com o botão direito do mouse e, na sequência selecionar a opção de análise PCA e o tipo de pré-processamento adequado, que neste caso é o “Autoescale”. O resultado será o gráfico dos scores apresentado na Figura 3b, e um vídeo com o passo a passo está incluído no Material Suplementar.

Pode-se observar a direita da Figura 3b a porcentagem de variância explicada em cada componente principal. Observa-se também que o software considerou até 6 PCs, isso porque temos apenas 6 linhas de informação quantitativa na tabela de dados numéricos utilizada na análise PCA.

Uma dúvida recorrente por parte dos discentes que cursaram a disciplina de quimiometria nos diferentes semestres foi a questão de interpretação do gráfico dos scores. Para tal, primeiramente, deve-se observar que a proposta inclui duas classes diferentes de aminoácidos conforme Figura 4: com grupos R aromáticos (quadro verde, Figura 4a), e com grupos R apolares (quadro azul, Figura 4c). A estrutura básica dos aminoácidos é apresentada na Figura 4b (quadro vermelho). Portanto, espera-se (ou procura-se) como resposta principal da análise PCA, no gráfico dos scores, uma separação dessas duas classes de aminoácidos, já que em função do grupo R tem-se diferentes propriedades dos aminoácidos de acordo com suas características de estrutura e tamanho.

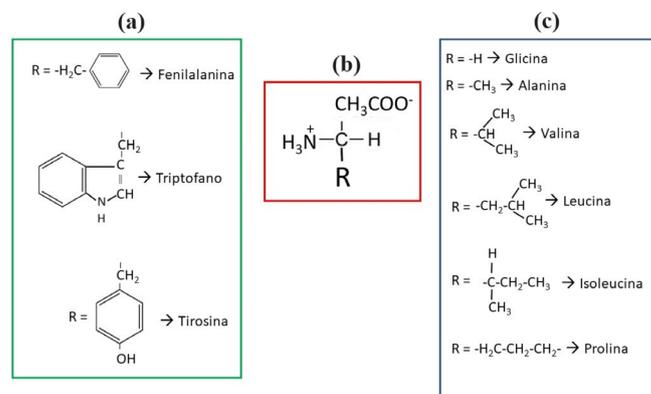


Figura 4. Estrutura química dos aminoácidos: grupos R aromáticos (a); estrutura básica de um aminoácido (b); grupo R apolar (c)

Para interpretar o gráfico de scores deve-se observar primeiramente a parte positiva de PC1, de onde pode-se concluir que os aminoácidos com grupos R aromáticos (fenilalanina, tirosina e triptofano) são separados dos demais aminoácidos considerados no estudo. Por outro lado, a parte negativa da PC1 separou a classe dos aminoácidos com grupos R apolares. Portanto, a primeira PC já foi responsável por responder nossa questão alvo, separando as classes de aminoácidos na parte positiva e negativa. Entretanto, podemos avaliar os resultados da PC2 ao verificar que dentre os aminoácidos considerados no estudo, a prolina se destaca na parte mais negativa desta PC, e podemos tentar esclarecer porque isso está acontecendo.

Para entender o porquê da observação da separação das classes de aminoácidos na PC1 recorre-se ao gráfico de loadings da PC1. Para tanto, deve-se selecionar o ícone “loadings”, indicado com a seta vermelha, que consta na Figura 3b. O resultado será o gráfico de loadings da PC1 apresentado na Figura 5a, onde se observa que na parte positiva de PC1 consta a variável Mr. Logo, somente essa variável é a responsável pela separação da classe de aminoácidos que se encontra em scores positivos da PC1, ou seja, os aminoácidos com grupos R aromáticos apresentam resultados diferenciados para essa variável. O próximo passo consiste em recorrer à tabela de dados numéricos, verificando que fenilalanina, tirosina e triptofano apresentam valores de Mr mais elevados que os demais aminoácidos.

Da mesma forma, os aminoácidos com grupos R apolares que se encontraram na parte negativa da PC1 no gráfico dos scores (Figura 3b), são separados devido às variáveis pK_1 , pK_2 , pI , IH

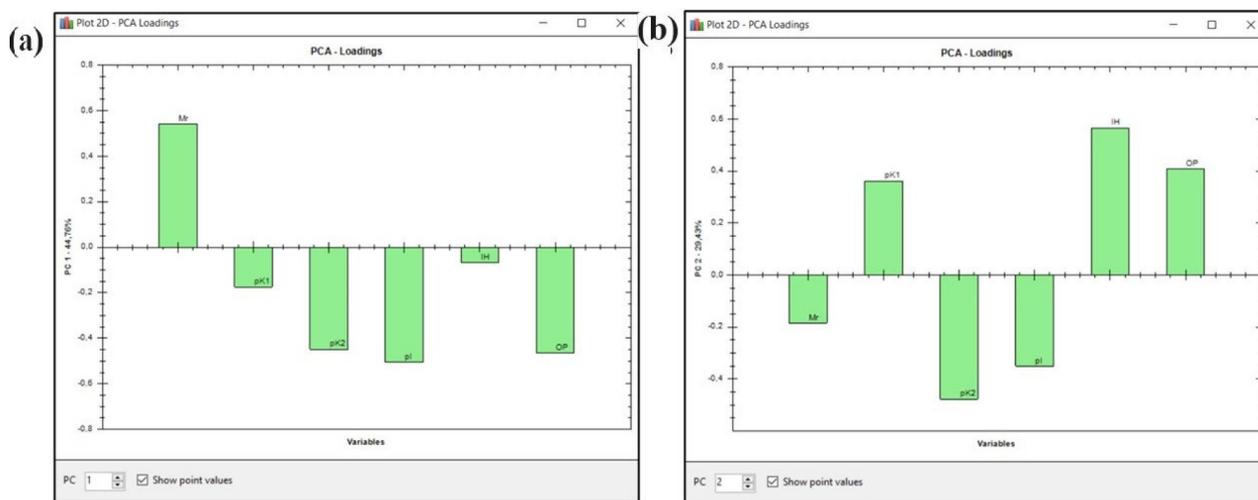


Figura 5. Gráficos de loadings: (a) loadings da PC1; (b) loadings da PC2

e OP (ocorrência nas proteínas). De posse dessa informação, e retomando aos dados numéricos da tabela, verifica-se que, em geral, os aminoácidos com grupos R apolares apresentam maiores resultados para pK_1 , pK_2 , pI, IH e OP.

Agora pode-se também buscar uma explicação para a distinção da prolina na parte mais negativa dos *scores* da PC2 (Figura 3b), avaliando o gráfico de *loadings* da PC2 (Figura 5b) que evidencia ser devido às variáveis Mr, pK_2 e pI. De fato, retomando a tabela de dados numéricos, verifica-se que a prolina apresenta maiores valores para pK_2 e pI, e um valor intermediário de Mr, em relação aos demais aminoácidos.

Este resultado pode ser explicado pois, o radical (R) da prolina e seu N α -amínico formam uma estrutura rígida em anel, com cinco átomos (Figura 6), conferindo a esse aminoácido características de pK_2 e pI que o difere dos demais. Além disso, a prolina, apresenta um grupo amino secundário, e não primário, como os demais aminoácidos do grupo R não polar.²⁶

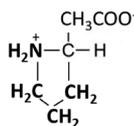


Figura 6. Estrutura química da prolina

É importante salientar que, neste caso, a PC1 (que apresenta a maior porcentagem de variância explicada) discriminou as duas classes diferentes de aminoácidos. Entretanto, algumas vezes, informações interessantes podem estar nas PCs com menores porcentagens de variância explicada.²⁷⁻²⁹ Nesta proposta, isso aconteceu na PC2 que discriminou a prolina dos demais aminoácidos e, a PC3 (que explica aproximadamente 15% da variância do conjunto de dados) se considerada, será observado que triptofano e tirosina são separados da fenilalanina devido ao caráter mais polar do triptofano e da tirosina quando comparados com a fenilalanina.

CONCLUSÕES

Neste trabalho, foi sugerido um banco de dados de propriedades de aminoácidos com grupos apolares e aromáticos como um exemplo simples e didático para uma atividade prática de análise de dados com PCA. Os resultados mostram-se acessíveis e de fácil compreensão para estudantes de química, da área de alimentos, farmacêutica ou áreas afins, pois a questão levantada é simples e exige apenas conhecimento básico de química.

O gráfico dos *scores* mostrou a separação entre as classes de aminoácidos, enquanto o gráfico de *loadings* permitiu explicar corretamente o porquê de a separação ter ocorrido, através da amplitude dos *loadings*.

Espera-se que esse instrumental possa favorecer uma melhor compreensão por parte dos estudantes sobre o PCA, bem como uma efetiva incorporação dos fenômenos que envolvem essa ferramenta e suas aplicações. Da mesma forma, espera-se que essa proposta didático-pedagógica possa subsidiar novas práticas docentes em disciplinas de quimiometria no ensino superior.

MATERIAL SUPLEMENTAR

O material suplementar desse trabalho está disponível em <http://quimicanova.sbq.org.br/>, na forma de arquivo PDF, e um vídeo com o passo a passo no formato mp4, com acesso livre.

AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - 306606/2020-8), a Gilson Augusto Helfer e alunos da disciplina QMM370 turma NQ7A de 2024/1 pelas discussões, levantamento de questões, e conclusões acerca do ponto flutuante e a UTFPR, Campo Mourão (Edital DIRPPG-CM 02/2024).

REFERÊNCIAS

- Kowalski, B. R.; *J. Chem. Inf. Model.* **1975**, *15*, 201. [Crossref]
- Kowalski, B. R.; *Anal. Chem.* **1980**, *52*, 112. [Crossref]
- Ferreira, M. M. C.; *Quimiometria: Conceitos, Métodos e Aplicações*, 1ª ed.; Editora da Unicamp: Campinas, 2015.
- de Barros Neto, B.; Scarmínio, I. S.; Bruns, R. E.; *Quim. Nova* **2006**, *29*, 1401. [Crossref]
- Wold, S.; Esbensen, K.; Geladi, P.; *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37. [Crossref]
- Guizzellini, T. A.; Rossi, A. V.; *XVI Encontro Nacional de Ensino de Química (XVI ENEQ) e X Encontro de Educação Química da Bahia (X Eduqui)*; Salvador, Brasil, 2012. [Link] acessado em Junho 2024
- Veras, G.; Alves, V. D.; Ramos, H. A.; Gomes, M.; Figueiredo, L.; Matias, E. V. S.; *Quim. Nova* **2022**, *45*, 1315. [Crossref]
- Lyra, W. S.; da Silva, E. C.; de Araújo, M. C. U.; Frago, W. D.; Veras, G.; *Quim. Nova* **2010**, *33*, 1594. [Crossref]
- Souza, A. M.; Poppi, R. J.; *Quim. Nova* **2012**, *35*, 223. [Crossref]

10. Pereira, F. M. V.; Pereira-Filho, E. R.; *Quim. Nova* **2018**, *41*, 1061. [Crossref]
11. Rios, C. A.; de Souza, A. M.; Poppi, R. J.; Breikreitz, M. C.; *Quim. Nova* **2022**, *45*, 1167. [Crossref]
12. Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O.; *Quim. Nova* **1999**, *22*, 724. [Crossref]
13. Guedes, T. L. M.; Soares, M. S. L.; das Neves, L. S.; de Lima, K. M. G.; *Quim. Nova* **2013**, *36*, 480. [Crossref]
14. Lima, K. M. G.; Trevisan, M. G.; Poppi, R. J.; de Andrade, C.; *Quim. Nova* **2008**, *31*, 700. [Crossref]
15. Breikreitz, M. C.; de Souza, A. M.; Poppi, R. J.; *Quim. Nova* **2014**, *37*, 564. [Crossref]
16. Teófilo, R. F.; Ferreira, M. M. C.; *Quim. Nova* **2006**, *29*, 338. [Crossref]
17. da Silva, K. C.; Caldeira, G. R. F.; Nogueira, K. G.; Canela, M. C.; Filgueiras, P. R.; Souza, M. O.; *Quim. Nova* **2023**, *46*, 98. [Crossref]
18. Valderrama, L.; Paiva, V. B.; Março, P. H.; Valderrama, P.; *Quim. Nova* **2016**, *39*, 245. [Crossref]
19. Lopes, E. R. C.; dos Santos, I. P.; Soares, F. L. F.; Pereira, E. F.; Braga, J. W. B.; *Anais do XV Encontro Nacional de Ensino de Química (XV ENEQ)*; Brasília, Brasil, 2010. [Link] acessado em Junho 2024
20. Saviani, D.; *Escola e Democracia: Teorias da Educação, Curvatura da Vara, Onze Teses sobre Educação e Política*, 44^a ed.; Autores Associados: Campinas, 2022.
21. Gasparin, J. L.; *Uma Didática para a Pedagogia Histórico-Crítica*, 5^a ed.; Autores Associados: Campinas, 2012.
22. Helfer, G. A.; Bock, F.; Marder, L.; Furtado, J. C.; da Costa, A. B.; Ferrão, M. F.; *Quim. Nova* **2015**, *38*, 575. [Crossref]
23. Nelson, D. L.; Cox, M. M.; *Princípios de Bioquímica de Lehninger*, 7^a ed.; Artmed: Porto Alegre, 2018.
24. Kyte, J.; Doolittle, R. F.; *J. Mol. Biol.* **1982**, *157*, 105. [Crossref]
25. Klapper, M. H.; *Biochem. Biophys. Res. Commun.* **1977**, *78*, 1018. [Crossref]
26. Ferrier, D. R.; *Bioquímica Ilustrada*, 7^a ed.; Artmed: Porto Alegre, 2019.
27. Moreira, I.; Scarminio, I. S.; *Talanta* **2013**, *107*, 416. [Crossref]
28. de Almeida, M. M. C.; Francisco, C. R. L.; de Oliveira, A.; de Campos, S. S.; Bilck, A. P.; Fuchs, R. H. B.; Gonçalves, O. H.; Valderrama, P.; Genena, A. K.; Leimann, F. V.; *Food Bioprocess Technol.* **2018**, *11*, 926. [Crossref]
29. de Almeida, M. M. C.; Francisco, C. R. L.; de Oliveira, A.; de Campos, S. S.; Bilck, A. P.; Fuchs, R. H. B.; Gonçalves, O. H.; Valderrama, P.; Genena, A. K.; Leimann, F. V.; *Food Bioprocess Technol.* **2018**, *11*, 940. [Crossref]