

14 - 01 | 2025

# APLICAÇÃO DA ANÁLISE DAS COMPONENTES PRINCIPAIS NA ANÁLISE DA INFORMAÇÃO NUTRICIONAL

## Application of Principal Component Analysis in the Analysis of Nutritional Information

## Aplicación del análisis de componentes principales en el análisis de información nutricional

Palmira Jorge Munguambe Condo<sup>1</sup>

<sup>1</sup> *Mestre em Gestão de Empresas pela USTM, Moçambique, <https://orcid.org/0009-0009-5506-3336>, [palmirajmunguambe@gmail.com](mailto:palmirajmunguambe@gmail.com).*

Autor para correspondência: [palmirajmunguambe@gmail.com](mailto:palmirajmunguambe@gmail.com)

*Data de recepção: 05-12-2024*

*Data de aceitação: 07-01-2025*

**Como citar este artigo:** Munguambe, P. J. (2025). Aplicação da Análise das Componentes Principais na Análise da Informação Nutricional. *ALBA - ISFIC Research and Science Journal*, 1(6), pp. 230-237. <https://alba.ac.mz/index.php/alba/issue/view/8>.

### RESUMO

O presente artigo tem como objectivo principal aplicar a Análise das Componentes Principais (ACP) na redução das variáveis associadas às quantidades das 300 pizzas dos dez diferentes tipos de informação nutricional disponibilizada por uma certa entidade. Para este efeito, utilizamos o software SPSS como suporte para a análise e tratamento de dados. Além disso, neste trabalho, apresentamos um quadro comparativo entre a Análise das Componentes Principais e a Análise Fatorial (AF). É importante destacar que, muitas vezes, essas técnicas são confundidas como se fossem as mesmas devido ao facto do software SPSS utilizar a mesma técnica na extração das componentes. Todas as hipóteses formuladas neste estudo usam como critério de decisão a rejeição da hipótese nula caso o valor-p seja inferior a 1% para a distância de Mahalanobis e 5% para os restantes testes.

**Palavras-chave:** Análise Fatorial, Análise das Componentes Principais, Nutrição.

### ABSTRACT

The main objective of this article is to apply Principal Component Analysis (PCA) to reduce the variables associated with the quantities of 300 pizzas of ten different types of nutritional information provided by a certain entity. To achieve this, we used the SPSS software as support for data analysis and processing. Additionally, in this study, we present a comparative framework between Principal Component Analysis and Factor Analysis (FA). It is important to note that these techniques are often confused as if they were the same because the SPSS software uses the same technique for component extraction. All hypotheses formulated in this study use the rejection of the null hypothesis as a decision criterion if the p-value is less than 1% for the Mahalanobis distance and 5% for the remaining tests.

**Keywords:** Factor Analysis, Principal Component Analysis, Nutrition.

### RESUMEN

El objetivo principal de este artículo es aplicar el Análisis de Componentes Principales (PCA) para reducir las variables asociadas a las cantidades de 300 pizzas de los diez tipos

diferentes de informação nutricional que põe a disposição uma determinada entidade. Para isso, utilizamos o software SPSS para respaldar o análise e o processamento de dados. Além disso, neste trabalho apresentamos um quadro comparativo entre o Análise de Componentes Principais e o Análise Factorial (AF). É importante ressaltar que muitas vezes se confundem estas técnicas como se fossem iguais devido a que o software SPSS utiliza a mesma técnica para extrair os componentes. Todas as hipóteses formuladas neste estudo utilizam o rejeito da hipótese nula como critério de decisão se o valor  $p$  é inferior a 1% para a distância de Mahalanobis e a 5% para os testes restantes.

**Palavras chave:** Análise Factorial, Análise de Componentes Principais, Nutrição.

### **Contribuição de autoria**

Concepção da ideia, pesquisa e revisão de literatura, aplicados informações resultantes dos instrumentos aplicados, análise estatística, preparação de tabelas, gráficos, preparação da base de dados, aconselhamento geral sobre o tema abordado, redação do original (primeira versão), revisão e versão final do artigo, correção do artigo, tradução de termos ou informações obtidas, revisão da aplicação do padrão bibliográfico aplicado.

## **INTRODUÇÃO**

O presente artigo tem como objetivo apresentar as diferenças existentes entre a Análise das Componentes Principais (ACP) e a Análise Factorial (AF). Para atingir este objetivo, recorreremos à literatura sobre as técnicas exploratórias fornecida e àquela pesquisada pelo estudante pesquisador. O problema abordado neste trabalho diz respeito à informação nutricional de 10 pizzas diferentes, recolhida numa amostra de 300 pizzas distribuídas em sete quantidades a enumerar: quantidade de água, proteínas, gorduras, cinzas, sódio, açúcar e quantidade de calorias por 100g.

Este artigo está estruturado em cinco capítulos principais. O primeiro capítulo corresponde à

introdução propriamente dita. No segundo capítulo, apresentamos os métodos utilizados na análise. O terceiro capítulo discute o caso em estudo, detalhando o problema da informação nutricional das pizzas. No quarto capítulo, apresentamos os resultados obtidos através da aplicação das técnicas de ACP e AF. Por fim, no quinto capítulo, apresentamos as conclusões do estudo, resumindo as principais diferenças entre as duas abordagens e destacando a relevância dos resultados para a análise da informação nutricional das pizzas

### *Análise das Componentes Principais*

Segundo Maroco (2007), a Análise das Componentes Principais (ACP) é uma técnica de análise exploratória multivariada que transforma um conjunto de variáveis correlacionadas em um conjunto menor de variáveis independentes. O autor salienta que essa técnica é um método de redução da complexidade dos dados. Assim como a Análise Factorial, a ACP tem por objetivo descrever os dados contidos em uma matriz de indivíduos e variáveis numéricas, em que  $p$  variáveis estão relacionadas com  $n$  indivíduos.

A ACP é considerada um método factorial porque a redução do número de variáveis não ocorre apenas por meio da seleção de algumas variáveis, mas também pela criação de novas variáveis sintéticas. Essas novas variáveis são obtidas por meio da combinação linear das variáveis originais usando os fatores. Para obter a Análise das Componentes Principais, é necessário formar combinações lineares das variáveis originais, representadas por  $Y_{n1} = x_{11}X_{n1} + x_{12}X_{n2} + \dots + x_{1p}X_{np}$ , onde  $x_{1p}$  são autovetores e  $X_{np}$  são os valores observados e originais.

### *Análise Factorial*

Maroco (2007) considera que a Análise Factorial (AF) é aplicada para identificar fatores em um determinado conjunto de medidas realizadas. É uma ferramenta que tenta reduzir um grande conjunto de variáveis para um conjunto mais significativo, representado pelos fatores. Esse método determina quais variáveis pertencem a quais

factores e o quanto cada variável explica cada factor.

Por outro lado, Pestana e Gageiro (2005) afirmam que a AF, assim como a Análise das Componentes Principais (ACP), descreve os dados contidos em uma matriz de indivíduos e variáveis numéricas, reduzindo-as: p variáveis estão relacionadas com n indivíduos. No entanto, eles acrescentam um elemento importante que diferencia a ACP da AF. Eles afirmam que a AF é mais ambiciosa porque postula um modelo que explica a correlação entre as variáveis observadas, pressupondo a existência de um número menor de variáveis não observáveis (factores) subjacentes aos dados. Esses fatores expressam o que é comum nas variáveis originais.

A realização da AF envolve quatro passos ou etapas principais, desde a entrada de dados até a rotação da matriz. O primeiro passo consiste em criar uma matriz das variáveis com n indivíduos. O segundo passo envolve a verificação do grau de correlação entre as variáveis, duas a duas. Maroco (2007) apresenta várias técnicas para a extração inicial dos factores, mas nesta pesquisa estamos interessados na técnica das componentes principais, pois é a mais usada na área de pesquisa.

Em relação à rotação dos factores, Maroco (2007) menciona a importância desse processo, pois nem sempre é possível atribuir significado empírico aos factores extraídos. Portanto, a rotação é necessária para obter um significado empírico para os factores extraídos. Os factores extraídos frequentemente apresentam dificuldades de interpretação, e a rotação ajuda a facilitar a interpretação desses factores.

Como pode ser percebido, o objectivo da extração de factores é encontrar um conjunto de factores que formem uma combinação linear das variáveis originais ou da matriz de correlações, assumindo que (X1, X2 ...Xn)

**Tabela 1- Valores de Kaiser- Meyer- Olkin (KMO)**

Valor de KMO	0.9-1.0	0.8-0.9	0.7-0.8	0.6-0.7	0.5-0.6	= < 0.5
Classificação da aplicação da AF	Excelente	Boa	Média	Mediocre	Mau mais ainda aceitável	Inaceitável

com correlações (C1j, C2j, ...Cnj), matematicamente pode-se expressar:

$$F_j = C_{1j}X_1 + C_{2j}X_2 + \dots + C_{nj}X_n$$

Maroco (2007) menciona que existem duas formas principais de procedimentos de rotação da matriz: a rotação ortogonal, conhecida como rotação *Varimax*, que mantém os factores não correlacionados entre si, e a rotação oblíqua, que permite que os factores se correlacionem entre si. O conceito essencial por trás da rotação de factores é identificar alguns factores que tenham variáveis com alta correlação e outros com variáveis que tenham baixa correlação. Nesta pesquisa, utilizaremos o método *Varimax*, que é uma técnica de rotação ortogonal, para melhor compreender a estrutura dos fatores e manter a independência entre eles.

*Pressupostos e Estatísticas na Aplicação de Análise Factorial*

*Estatísticas de Kaiser – Meyer Olkin*

Antes de aplicar a Análise Factorial (AF), vários pressupostos precisam ser estabelecidos. Esses pressupostos incluem a realização de testes de normalidade tanto univariada quanto multivariada, bem como a avaliação da qualidade dos dados por meio da medida de Kaiser-Meyer-Olkin (KMO).

A medida de adequação de Kaiser-Meyer-Olkin é um método utilizado na análise multivariada para avaliar a qualidade dos dados. Conforme Kaiser e Rice (1974), citados por Maroco (2007), o KMO é uma medida da homogeneidade das variáveis que compara as correlações simples com as correlações parciais observadas entre as variáveis.

A fórmula para calcular essa estatística é a seguinte:

$$KMO = \frac{\sum_{j=1}^p \sum_{m=1, m \neq j}^p r_{jm}^2}{\sum_{j=1}^p \sum_{m=1, m \neq j}^p r_{jm}^2 + \sum_{j=1}^p \sum_{m=1, m \neq j}^p r_{pj}^2} \quad (Eq.1).$$

### Estatística de Esfericidade de Bartlett

O segundo pressuposto é o da Esfericidade de Bartlett, que se baseia na distribuição estatística qui-quadrado e testa a hipótese nula (H0) de que a matriz de correlação é uma matriz identidade, ou seja, que não há correlação entre as variáveis. Conforme Maroco (2007), o teste da Esfericidade de Bartlett é formulado da seguinte maneira:

H0:  $\pi = I$  (onde  $\pi$  é a matriz das correlações e  $I$  é a matriz de identidade)

Ha:  $\pi \neq I$  (A matriz  $I$  é diferente da matriz de identidade)

Resumidamente, as hipóteses são as seguintes:

H0:  $|R| = 1$  ou  $|R| = I$  (a matriz das correlações é igual à matriz identidade)

Ha:  $|R| \neq 1$  ou  $|R| \neq I$  (a matriz das correlações é diferente da matriz identidade)

No caso de não serem cumpridas essas duas premissas principais, será necessário eliminar as variáveis cuja medida de adequação da amostra seja inferior a 0,5, conforme indicado por Pestana e Gageiro (2005). A eliminação é baseada na matriz anti-imagem, em que as variáveis com valores de Medida de Adequação da Amostra (MAS) inferiores a 0,5 são removidas.

A estatística do teste é calculada da seguinte forma:

$$X^2 = -\ln|R| \left[ (n-1) - \frac{2p+5}{6} \right] \text{ (Eq.2)}$$

Esta estatística tem uma distribuição qui-quadrado com graus de liberdade

$$g = \frac{p(p-1)}{2} \text{ (Eq.3)}$$

Onde:  $n$  é o tamanho de amostra;  $p$  é o número de variáveis e  $|R|$  é o determinante da matriz das correlações.

### Matriz de Correlação

A matriz de correlação exibe as correlações simples, representadas por "r", entre todos os possíveis pares de variáveis incluídas na análise. Os elementos da diagonal principal são todos iguais a 1.

### Comunalidade

Segundo Maroco (2007), a comunalidade refere-se à parcela ou percentagem da variância que uma variável compartilha com todas as outras variáveis consideradas. É a proporção da variância que é explicada pelos factores comuns. Para diferenciar a técnica da Análise das Componentes Principais (ACP) da Análise Fatorial (AF), preferimos apresentar um quadro comparativo entre as duas análises, resumindo da seguinte forma:

**Tabela 2- Diferenças entre ACP e AF**

<b>Análise das Componentes Principais</b>	<b>Análise Factorial</b>
As componentes principais são definidas como combinações lineares das variáveis originais.	As variáveis originais são expressas como combinações lineares dos factores.
Procura explicar a maior parte da variância total existente nas variáveis.	Procura explicar a co-variância e/ou correlações entre as variáveis.
Não requer nenhuma suposição	Faz várias suposições-chave
As componentes principais são únicas, assumindo distintos autovalores	Está sujeita a uma rotação arbitrária (se alterar o número de factores, os factores estimados podem mudar)
É uma manipulação matemática	Uma técnica estatística.

## MATERIAIS E MÉTODOS

Para a elaboração deste artigo, recorreremos à revisão bibliográfica para embasar teoricamente a aplicação da Análise das

Componentes Principais (ACP). No que diz respeito ao tratamento e análise dos dados, utilizamos o Software Estatístico para Ciências Sociais (SPSS) versão 25. Para descrever os dados relacionados à informação nutricional, aplicamos a análise estatística descritiva.

Na análise descritiva, priorizamos a identificação de outliers por meio da distância de Mahalanobis, a abordagem de valores ausentes, bem como o cálculo da média, desvio padrão, coeficiente de variação, assimetria e curtose da distribuição dos dados de informação nutricional.

As principais técnicas empregadas neste artigo incluem a Análise das Componentes Principais (ACP), utilizando o método de rotação Varimax, para analisar sete variáveis relacionadas à informação nutricional. No entanto, antes de aplicar essa técnica, realizamos uma série de verificações e considerações, como a identificação de valores ausentes, a detecção de valores discrepantes em cada variável, análise das correlações entre as variáveis, avaliação do critério de adequação da amostra utilizando o índice de Kaiser-Meyer-Olkin (KMO) e realização do teste de esfericidade de Bartlett.

Os dados sobre a informação nutricional foram obtidos de fontes secundárias e se referem às quantidades de água, proteínas, gorduras, cinzas, sódio, açúcar e calorias por 100g. Como ponto de partida, formulamos as seguintes perguntas: Quais são as principais quantidades dos ingredientes que compõem a pizza em termos de nutrição para os indivíduos? Quais ingredientes contribuem de forma positiva ou negativa para a composição nutricional da pizza?

## RESULTADOS E DISCUSSÃO

Nesta secção, realizamos uma análise descritiva dos dados, verificando valores discrepantes e a quantidade de não respostas para as seguintes variáveis de informação nutricional por 100g: quantidade de água, quantidade de proteínas, quantidade de gorduras, quantidade de cinzas, quantidade de sódio, quantidade de açúcar e calorias.

Ao examinar a Tabela 3, que apresenta as variáveis com valores discrepantes em comparação com os demais valores, bem como estatísticas descritivas e distribuição de frequência dos dados, observamos que não foram identificados valores em falta (omissos) para nenhuma das variáveis. No entanto, notamos a presença de *outliers* (valores

discrepantes) para todas as variáveis, com exceção da variável "açúcar". As demais variáveis apresentam valores discrepantes mais significativos quando comparados com os demais tipos de pizza.

Adicionalmente, realizamos uma análise da variabilidade dessas quantidades por tipo de pizza. Observamos que as variáveis "Quantidade de água por 100g" e "Calorias por 100g" exibem uma variabilidade moderada em torno da média, uma vez que seus coeficientes de variação estão acima de 15% e abaixo de 30%. Por outro lado, as demais variáveis demonstram uma grande variabilidade em relação à média. Quanto à assimetria e curtose, podemos observar que essas variáveis apresentam uma distribuição assimétrica, com alguns valores muito inferiores à média (assimetria positiva) e alguns valores muito superiores à média (assimetria negativa). Em relação à curtose, essas variáveis exibem diferentes graus de achatamento da distribuição, com algumas sendo platicúrticas e outras leptocúrticas.

Considerando o tipo de dados em análise, acreditamos que a aplicação da Análise das Componentes Principais (ACP) é ideal por várias razões. Primeiro, o tamanho da amostra é proporcional ao número mínimo necessário de casos por variável, com uma razão de aproximadamente 43 casos por variável, o que é adequado. Segundo, não há valores ausentes para nenhuma das variáveis, o que não afecta a análise. Além disso, os valores discrepantes para as categorias de cada tipo de pizza são estatisticamente insignificantes, com exceção do outlier com o ID 24069, que é estatisticamente significativo, com uma probabilidade (valor p) de 0.1% calculada pela distância de Mahalanobis.

Para validar a adequação da aplicação da ACP, consideramos vários testes e critérios. Primeiro, aplicamos o critério de Kaiser-Meyer-Olkin (KMO), que é uma medida de adequação da amostra. Verificamos que, ao trabalhar com todas as variáveis em questão sobre a informação nutricional das 300 pizzas de 10 tipos diferentes, a ACP não é aplicável, uma vez que o valor do KMO é de apenas 0.445, abaixo do limiar mínimo de 0.5. Isso significa que a aplicação da ACP não é

aceitável, apesar de a variância total explicada ser de 92.32%, sendo que 55.70% pertence à primeira componente e a segunda componente contribui com 36.62% na variância total explicada (vide Apêndice). Como medida corretiva para melhorar o grau de ajustamento da adequação da amostra, recorreremos à matriz anti-imagem para remover as variáveis com correlações abaixo de 0.5. As variáveis "Quantidade de proteínas por 100g" e "Calorias por 100g" foram retiradas, resultando em um KMO aceitável de 0.705, que se enquadra nos intervalos recomendados por Maroco como uma boa medida de adequação da amostra.

Além disso, realizamos o Teste de Esfericidade de Bartlett, que indicou que a matriz de correlações da população é significativamente diferente da matriz identidade, uma vez que a hipótese nula estabelecida antes do teste foi rejeitada, com um valor p (Sig.) de 0.0001,

inferior a 0.05. Com todas essas condições atendidas, avançamos com a aplicação da ACP nas cinco variáveis restantes associadas às quantidades de dados de informação nutricional de 10 tipos diferentes de pizzas.

Para reter as componentes principais, utilizamos o critério da Análise das Componentes Principais (ACP). A proporção da variância dos dados explicada por cada componente foi de 78.68%. Cada variável contribuiu significativamente para a componente retida. Essa componente foi denominada "Carboidratos", uma vez que todas as variáveis em análise estão relacionadas às quantidades de carboidratos.

Se todas as sete variáveis fossem considerados, teríamos duas componentes principais: "Quantidade de alimentos nutricionais protetores" e "Quantidade de alimentos nutricionais energéticos" para a primeira e segunda componentes, respetivamente.

**Tabela 3-Análise de Estatística descritiva por tipo de pizza**

Código da Variável	Valores ausentes e valores discrepantes dos restantes valores		Estatísticas descritivas e distribuição da frequência				
	Valores ausentes (%)	Variáveis com valores discrepantes dos restantes valores por tipo de Pizza	Média	Desvio padrão	Coefficiente de variação (%)	Assimetria	Kurtose
agua	0	Cheeseham (1 grave), 3queijos (1 grave), vegetariana (3 graves), quatro estações	40.90	9.55	23.36	-0.04	-1.53
proteina	0	Cheeseham (2 graves), Havaiana (1 grave)	13.37	6.43	48.11	0.87	-0.75
gordura	0	Vegetariana (1 grave)	20.23	8.98	44.37	1.56	1.91
cinza	0	Calzone (1 grave), Vegetariana (1 grave)	2.63	1.27	48.22	0.68	-0.88
sodio	0	Cheeseham (2 graves), Calzone (1 grave)	0.67	0.37	55.33	1.82	2.26
acucar	0	Todos são suaves	22.86	18.03	78.85	0.05	-1.67
calorias	0	3queijos (2 grave), Vegetariana (1 grave)	3.27	0.62	18.96	1.07	1.19

**Tabela 4: Medida de adequação da amostra KMO e teste de esfericidade de Bartlett's**

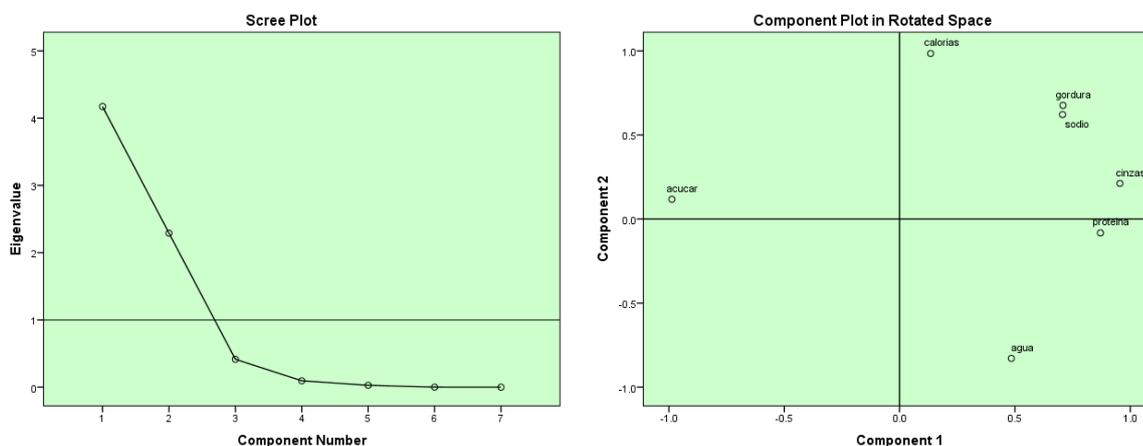
KMO and Bartlett's Test			
		Todas variáveis	5 Variáveis
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.455	.705
Bartlett's Test of Sphericity	Approx. Chi-Square	7135.057	2072.905
	df	21	10
	Sig.	0.000	0.000

**Tabela 5: Proporção de variância explicada e cargas factoriais da matriz das componentes**

Variáveis	Comunalidades	Cargas factoriais
Quantidade de proteínas por 100 g	.659	<b>.812</b>
Quantidade de gordura por 100 g	.761	<b>.872</b>
Quantidade de cinzas por 100 g	.956	<b>.978</b>
Quantidade de sódio por 100 g	.735	<b>.857</b>
Quantidade de açúcares por 100 g	.823	-.907
Variância total explicada (%)	78.68	

**Método de extração:** Análise das componentes Principais e rotação: *Varimax*

Variáveis	Componentes		Communalidades
	1	2	
Quantidade de água por 100 g		-.951	.922
Quantidade de proteínas por 100 g	.774		.765
Quantidade de gordura por 100 g	.912		.958
Quantidade de cinzas por 100 g	.964		.957
Quantidade de sódio por 100 g	.890		.885
Quantidade de açúcares por 100 g	-.868		.988
Calorias por 100 g		.859	.987
Variância total explicada	55.70%	36.62%	92.32%



**Figura 1:** Scree plot e plotagem da componente num espaço rotacionado

## CONCLUSÃO

No âmbito da distinção e aplicação da Análise das Componentes Principais (ACP) e da Análise Factorial (AF), concluímos que existem diferenças ligeiras, o que muitas vezes leva à confusão entre esses dois métodos. A ACP tem como objectivo principal determinar um conjunto de componentes, em que cada

componente é uma combinação das variáveis originais. Por outro lado, a técnica da AF procura determinar um conjunto de factores, em que cada variável original é uma combinação desses factores. Portanto, é importante destacar que a ACP não é a mesma coisa que a Análise Factorial.

No contexto da aplicação da ACP em uma amostra de 300 pizzas de 10 tipos diferentes,

relacionadas à informação nutricional, concluímos que, das sete variáveis disponíveis, conseguimos reduzir para dois indicadores principais, que chamamos de componentes. Além disso, observamos que a medida de adequação da amostra (KMO) era inicialmente inaceitável. No entanto, identificamos que era possível aplicar algumas medidas corretivas para melhorar o ajustamento da medida, como a utilização da matriz anti-imagem e a eliminação das variáveis cujos coeficientes na diagonal principal eram inferiores a 0.5. A variância total explicada pelo modelo, mesmo desconsiderando os pressupostos não atendidos, foi de 92.32%, sendo que a primeira componente, por si só, conseguiu explicar mais da metade (55.70%) dessa variância total.

Os principais indicadores da informação nutricional identificados foram as quantidades de alimentos protectores e energéticos. Uma conclusão relevante deste estudo é enfatizar a importância do trabalho e até mesmo sugerir possíveis aplicações ou extensões futuras. Além disso, vale ressaltar que a ACP é uma técnica que não requer suposições rigorosas e é uma das abordagens úteis para reduzir ou manipular um conjunto de variáveis, criando combinações lineares significativas.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Pestana, M. H e. Gageiro, J. N. (2005). *Análise de Dados para Ciências sociais: A Complementaridade do SPSS*, 4ª Edição, Edições Sílabo Lda, Lisboa – Portugal.
- Maroco, J. (2007). *Análise. Estatística com utilização do SPSS*. 3ª edição, edições Sílabo, Lisboa, Portugal.
- Reis, E. (2001). *Estatística Multivariada Aplicada*, 2ª Edição, Edições Sílabo, Lisboa.
- <https://statistics.laerd.com/spss-tutorials/principal-components-analysis-pca-using-spss-statistics.php>
- <https://stats.idre.ucla.edu/spss/seminars/efa-spss/>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and

recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202.

<https://doi.org/10.1098/rsta.2015.0202>

Shlens, J. (2014). A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100.

<https://arxiv.org/abs/1404.1100>

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459. <https://doi.org/10.1002/wics.101>

Ringnér, M. (2008). What is principal component analysis? *Nature Biotechnology*, 26(3), 303–304. <https://doi.org/10.1038/nbt0308-303>

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer. <https://doi.org/10.1007/b98835>.

*Munguambe, P. J. (2025). Aplicação da Análise das Componentes Principais na Análise da Informação Nutricional.*