Análise de componentes principais em dados agrícolas de produção de soja

Principal component analysis of agricultural data on soybean production

[1]Rafael Queiroz, [2]Fabiane Silva, [3]Kuang Hongyu

[1]Graduando em Estatística – Universidade Federal de Mato Grosso
[2]Professora Doutora; Departamento de Estatística – Universidade Federal de Mato Grosso
[3]Professor Doutor; Departamento de Estatística – Universidade Federal de Mato Grosso

RESUMO: O agronegócio desempenha um papel fundamental na economia brasileira, contribuindo significativamente para o crescimento do país. Este artigo tem como objetivo realizar uma análise de dados do agronegócio utilizando a técnica de análise de componentes principais por meio da matriz de correlação, validando sua relevância para o setor e apresentar as possibilidades que o uso de técnicas estatísticas trás para a maximização de produção de soja na lavoura. As variáveis selecionadas para análise incluem: área, dose por hectare, quantidade consumo, área execução, dose real, valor, vazão, peso bruto, impureza e umidade, a técnica de análise de componentes principais foi aplicada para reduzir a dimensionalidade dos dados, identificar padrões subjacentes e investigar as relações entre as variáveis selecionadas. O critério de Kaiser foi utilizado para validar a adequação dos dados à análise de componentes principais, considerando os autovalores das variáveis para determinar o número de componentes principais significativos a serem retidos. Dessa forma, conclui-se que a análise de componentes principais além de ser uma ótima forma de economizar recursos computacionais minimizando o volume de variáveis a serem analisadas, também explica de forma satisfatória o comportamento dinâmico das informações, provando a relevância da técnica para compreensão das peculiaridades da produção de soja na agricultura e fornecer oportunidades valiosas para tomada de decisão estratégica.

Palavras Chave: Analise de componentes principais. Matriz de correlação. Técnicas estatísticas.

***ABSTRACT*:** Agribusiness plays a fundamental role in the Brazilian economy and has significantly contributed to the country's growth. This article aims to analyze agribusiness data using the principal component analysis technique, validating the relevance to the sector and presenting the possibilities the statistical techniques use bring for maximizing soybean production in the field. The selected variables for analysis include: area, dose per hectare, quantidade consumo, área execução, dose real, valor, vazão, peso bruto, impureza e umidade. The principal component analysis technique was applied to reduce data dimensionality, identify underlying patterns, and investigate relationships among the selected variables. The Kaiser criterion was used to validate the suitability of the data for principal component analysis, considering the variables' eigenvalues to determine the number of significant principal components to be retained. Because of that, it is concluded that principal component analysis, besides being an excellent way to save computational resources by minimizing the volume of variables to be analyzed, also adequately explains the dynamic behavior of information, proving the technique's relevance for understanding the peculiarities of soybean production in agriculture and providing valuable insights for strategic decision-making.

Keywords: Principal component analysis. Correlation matrix. Statistical techniques.

INTRODUÇÃO

O agronegócio é um setor vital para a economia de muitos países, responsável por garantir a segurança alimentar e gerar empregos em diferentes regiões (CAMPEÃO, 2020). Com a crescente adoção de tecnologias avançadas e o aumento do uso de dados no campo, a análise de dados tornou-se uma ferramenta importante para auxiliar os profissionais do agronegócio a tomarem decisões mais informadas e assertivas. A análise multivariada é uma técnica estatística utilizada para analisar conjuntos de dados

E&S - Engineering and Science ISSN: 2358-5390 DOI: 10.18607/ES20231215788 Volume 12, Número 3

com múltiplas variáveis simultaneamente, buscando identificar as relações e padrões existentes entre elas (HAIR, 2009). Uma das técnicas mais utilizadas na análise multivariada é a análise de componentes principais (ACPP), que permite reduzir a dimensionalidade dos dados, mantendo a maior parte da variação original (HONGYU, 2015).

A análise de componentes principais pode ser uma ferramenta poderosa para auxiliar no entendimento dos dados do agronegócio. Ao aplicar a ACP em um conjunto de dados do agronegócio, é possível identificar as principais fontes de variação, reduzir a dimensionalidade dos dados e obter uma visualização mais clara e simplificada das relações entre as variáveis (FILHO, CAMPOS, LEMOS, 2023).

A utilização da matriz de correlação em conjunto com a PCA permite identificar as variáveis mais relevantes para a análise e obter uma visualização dos dados em um espaço bidimensional ou tridimensional, facilitando a interpretação dos resultados. Além disso, a análise de componentes principais pode ser aplicada em diferentes áreas do agronegócio, desde o monitoramento de culturas até a análise de dados de mercado.

Na análise de dados de produtividade de culturas, a ACP pode ajudar a identificar quais fatores influenciam mais na produção, como clima, fertilidade do solo e uso de fertilizantes, e a identificar quais variáveis são mais importantes para o sucesso da produção. Já na análise de dados de mercado, a ACP pode ser utilizada para identificar quais variáveis estão mais correlacionadas com o preço de determinado produto, como a oferta, a demanda e os custos de produção (HANKE, 2022).

Dessa forma, a aplicação da ACP no agronegócio pode fornecer insights valiosos e ajudar os profissionais do setor a tomar decisões mais informadas e estratégicas. O presente artigo, portanto, tem como objetivo apresentar uma visão geral da análise multivariada e da análise de componentes principais, com enfoque em sua aplicação no agronegócio e seus potenciais benefícios para a tomada de decisão nesse setor.

MATERIAIS E MÉTODO

A área de estudo compreende uma fazenda próxima a cidade de sapezal, localizado na região Centro-Oeste do Brasil. Esta fazenda em específico é voltada para a maximização da produção de soja, portanto todos seus equipamentos de plantio, colheita, infraestrutura de tratamento de sementes e armazenamento de insumos são calibradores para garantir a máxima performance na produção da cultura. A escolha do plantio é meticulosa, envolve a disponibilidade de água na região, fertilidade e características do solo, e a topografia da região.

Outra característica importante sobre a área dessa fazenda é que todo o campo é preparado com cultivo sustentável, isso significa que toda a área é tratada de forma a manter sua característica original, garantindo a preservação ambiental da região. Tudo isso com o intuito de maximizar a produção de soja.

O conjunto de dados engloba variáveis de produção, área, tratamento do solo e tratamento da cultura. Essas variáveis são: X_1: Área Prevista; X_2: Dose de Fertilizante Por Hectare; X_3: Quantidade Consumo de Água (Aplicação); X_4: Area Execução; X_5: Dose Real de Insumo; X_6: Valor de Produção; X_7: Vazão; X_8: Peso Bruto do Grão; X_9: Impureza do Grão; X_{10}: Umidade do Grão. A obtenção dos componentes é implementada através da análise diagonal de matrizes simétricas positivas. Portanto, é

possível calcular os componentes principais de forma assertiva e utilizados de acordo com a necessidade proposta.

Dentre todas as possibilidades mapeadas para aplicação do ACP, destaca-se a capacidade de resolver problemas de multicolinearidade em regressões lineares, estimar fatores na análise fatorial, outra técnica de análise multivariada bem difundida entre a comunidade cientifica, realizar a modelagem de interação entre fatores em experimentos sem repetição, estudos de divergências e agrupamento entre variáveis que possuem mesma características etimológicas ou de função, entre outras possibilidades (HONGYU, 2012; JOHNSON; WICHERN 1998).

A ACP tem como principal característica a retirada de multicolinearidade nas variáveis, isso permite transformação variáveis que anteriormente seriam analisadas de forma individual em um conjunto de variáveis interrelacionadas, ou seja, um novo conjunto de informação não correlacionada, chamado de componente principal. Além disso, reduz o volume de variáveis a eixos de representação. Esse eixo é perpendicular, o que explica a variação dos dados de forma independente (HONGYU, 2015).

Assim como todo modelo estatístico, a ACP possui desvantagens e sensibilidades que devem ser levadas em consideração na sua implementação. Entre elas a sensibilidade a outliers, o fato de não ser bem implementada em dados com dupla ausência, valores nulos ou ausentes, e quando se tem um fluxo de variáveis maior que o volume de dados amostrais analisado. Pois conforme explicado por HONGYU (2015), ao reduzir o número de variáveis há uma perda significativa de informação de variabilidade. Logo é necessário que a parte explicada seja o padrão de resposta e a outra parte seja o ruído, ou seja, erro de medida e redundância. A ACP não será eficiente em casos cuja variável original é pouco correlacionada, com o caso de **R** = **I**, os componentes principais nesse cenário não são as próprias variáveis originais.

Sejam as variáveis X_1, X_2, ..., X_p em cada um de *n* indivíduos ou unidade experimental, esse conjunto de *n x p* medida original de uma matriz de dados **X** *(n x p)*:

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

Para obtenção dos componentes principais, seja um conjunto de *p* variáveis X_1, X_2, ..., X_p com média μ_1, μ_2, ..., μ_p e variância σ_1, σ_2, ..., σ_p, respectivamente. Essas variáveis não são independentes e, portanto, possuem covariância entre a *i-éssima* e *k-éssima* variável definida por σ_{jk}, para $i \neq l = 1, 2, ..., p.$ Então as *p* variáveis pode ser expressa na forma vetorial por: **X** = $[X_1, X_2, \ldots, X_P]$, com vetor de média **μ** = [μ_1, μ_2, ..., μ_p] e matriz de covariância:

$$\sum = \begin{pmatrix} \sigma_{11}^2 & \cdots & \sigma_{1p}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \cdots & \sigma_{np}^2 \end{pmatrix}$$

Encontram-se nos pares de autovalores e autovetores (λ_1, e_1), (λ_2, e_2), ..., (λ_P, e_P) em que $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$, associados a $\sum$ e então ao *i-éssimo* componente principal. (JOHNSON; WICHERN, 1998; HONGYU, 2015):

$$\boldsymbol{Z}_i = \boldsymbol{e}_i^1 \boldsymbol{X} = e_{i1}X_1 + e_{i1}X_2 + \ldots + e_{ip}X_p$$

A variável $\boldsymbol{Z}_i$ é uma variável latente, ou seja, não é mensurada a partir do experimento ou levantamento amostral. (JOHNSON; WICHERN, 1998). Com isso será possível determiná-lo através de *p* variáveis contidas no vetor $\boldsymbol{X}$. Assim, será possível maximizar a variabilidade da variável latente $\boldsymbol{Z}_j$. A variância $\boldsymbol{Z}_j$ é dada por:

$$Var(\boldsymbol{Z}_i) = Var(\boldsymbol{e}_i'\mathbf{X}) = \boldsymbol{e}_i' Var(\boldsymbol{X})\boldsymbol{e}_i = \boldsymbol{e}_i' \sum \boldsymbol{e}_i$$

Em que i = 1, ..., p.

Utilizando a decomposição espectral da matriz ∑, DADA POR ∑ = ***PAP'***, em que $\boldsymbol{P}$ é a matriz composta pelos autovetores de ∑ em suas colunas e $\boldsymbol{A}$ é a matriz diagonal de autovalores de ∑, então, tem-se que:

$$tr = tr(\boldsymbol{PAP'}) = tr(\boldsymbol{AP'P}) = tr(\boldsymbol{AI}) = tr(\boldsymbol{A}) = \sum_{i=1}^{p} \lambda_i$$

$$\boldsymbol{A} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}$$

Etr(∑) é dada pela soma dos elementos diagonais:

$$Etr(\sum) = \sum_{i=1}^{p} \sigma_{ii} = \sum_{i=1}^{p} \lambda_i$$

Portanto, a variabilidade total contida nas variáveis originais é igual a variabilidade total contida nos componentes principais (JOHNSON; WICHERN, 1998). A contribuição de cada componente principal $\boldsymbol{Z}_i$ é expressa em porcentagem, e a explicação individual de cada componente pode ser calculada, por exemplo, para *k-ésimo* componente principal a proporção da explicação é dada por:

$$C_k = \frac{Var(\boldsymbol{Z}_i)}{\sum_{i=1}^{p} Var(\boldsymbol{Z}_i)} .100 = \frac{\lambda_i}{\sum_{i=1}^{p} \boldsymbol{\lambda}_i} .100 = \frac{\lambda_i}{Tr(\boldsymbol{S})} .100$$

Pela proporção de explicação da variância total, que o modelo de *k* componentes principais é responsável, podemos determinar o número de componentes que se deve obter. Em muitos casos, adota-se modelos que explique 70% da variação total ou mais (JOHNSON; WICHERN, 1998).

Em geral escolhe-se o componente principal de maior importância (o primeiro componente principal) como sendo aquele de maior variância, que explique o máximo de variabilidade dos dados, o segundo componente é o que apresenta a segunda maior variância e assim sucessivamente (HONGYU, 2015).

Por outro lado, os últimos componentes principais serão responsáveis por direções que não estão associadas a muita variabilidade. Em outras palavras, esses últimos componentes principais identificarão a relação linear entre as variáveis originais

próximas da constante. (JOHNSON; WICHERN, 1998; FERREIRA, 2011; HONGYU, 2015).

Por fim, o critério de Kaiser (KAISER, 1958) permite selecionar os componentes principais (CPs) que expliquem a maior parte da variação dos dados. Este critério obtém CPs com valores próprios maiores do que a unidade ($\lambda_i > 1$), isto é, os principais componentes que expliquem a maior parte da variação no conjunto de dados (SAVEGNAGO, 2011; HONGYU, 2015).

Para identificação dos indivíduos, temos que o código pode variar de um a trinta e quatro, sendo essa as subáreas da unidade em que os dados foram extraídos.

O software utilizado para realização do estudo foi o R versão 4.3.0, uma linguagem de programação e ambiente de desenvolvimento estatístico utilizado para análise de dados. Para manipular os dados foi utilizado o pacote "*readxl*", que permite ler e carregar dados de planilhas de forma direta. Além disso, utilizou-se o pacote "*FactorMineR*" e "*psych*", que fornecem uma ampla variedade de aplicações multivariadas incluindo análise de componentes principais.

RESULTADOS/ DISCUSSÕES

Ao analisar os dados com a técnica de análise dos componentes principais, os respetivos autovalores e porcentagens da variância explicada por cada um estão na Tabela 1. Os dois primeiros PCs foram responsáveis por 83,34% da variação total acima das variáveis de produção, armazenamento e área da fazenda, em que o PC1 foi responsável por 65,96% o segundo PC2 por 17,38% das variações dos dados.

Logo para determinação do número de componentes principais verificou-se que como os dois primeiros CPs gerados a partir desta análise que tem autovalores *> 1 ($\lambda i > 1$)* (KAISER, 1958; FRAGA, 2016) e foi responsável por 83,34% da variância total no conjunto de dados, os quatro CPs foram retidos, com o auxílio do screeplot (Figura 1). Portanto, os dois primeiros componentes principais resumem efetivamente a variância amostral total e podem ser utilizados para o estudo do conjunto de dados.

Em estudo com 11 características de produção com aves de corte, verificaram que apenas três componentes principais eram suficientes para explicar 77% da variância total das características. Por sua vez, Meira et al. (2013), apresentou 13 características morfofuncionais de cavalos da raça Mangalarga Machador, onde foi identificado 6 componentes principais com autovalores inferiores a 0,7, os quais explicaram 78,57% da variação total das informações (FRAGA, 2016).

Já de acordo com Hongyu (2015), os dois primeiros PCs foram responsáveis por 68.13% da variação total, sobre a taxa de criminalidade de algumas cidades dos Estados Unidos, em que o PC1 foi responsável por 49,37% e o segundo, PC2, por 18,76% das variações dos dados.

Para a determinação do número de componentes principais utilizou-se os dois primeiros CPs gerados a partir desta análise que tem autovalores *> 1 ($\lambda i > 1$)* (KAISER, 1958; FRAGA, 2016).

Tabela 1 – Componentes principais (CPs), autovalores (λ_i) e porcentagem da variância explicada e proporção acumulada (%) pelos componentes.

Componente Principal	Autovalores	Proporção	Proporção Acumulada (%)
PC1	6,60	65,96	65,96
PC2	1,74	17,38	83,34
PC3	0,65	6,45	89,79
PC4	0,41	4,07	93,86
PC5	0,31	3,11	96,97
PC6	0,16	1,62	98,60
PC7	0,07	0,71	99,31
PC8	0,05	0,49	99,80
PC9	0,02	0,20	100,00
PC10	0,00	0,00	100,00

Figura 1 – O screeplot dos autovalores dos componentes principais.

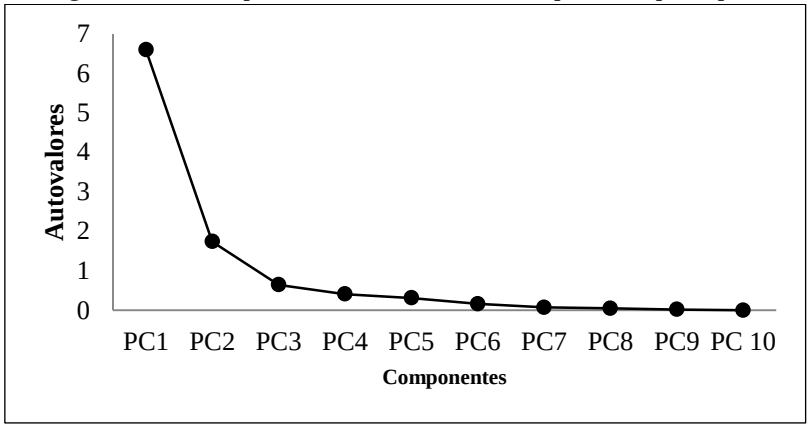


Tabela 2 – Coeficientes de ponderação das características com os dois primeiros componentes.

Variáveis	CP1	CP2
Área	0,95	0,40
Dose Por Hectare	0,90	-0,23
Quantidade Consumo	0,93	0,18
Área Execução	0,96	-0,01
Dose Real	0,89	-0,23
Valor	0,64	-0,65
Vazão	0,73	-0,78
Peso Bruto	0,79	0,44
Impureza	0,55	0,69
Umidade	0,65	0,46

A escolha das variáveis para análise no agronegócio foi baseada em critérios como relevância, disponibilidade dos dados e importância para a compreensão do fenômeno em estudo. Além disso, com essas variáveis foi possível entender aspectos produtivos, econômicos e até fatores relacionados à qualidade e sustentabilidade.

Como intuito de se entender a importância de cada variável na construção dos dois componentes foi calculado a correlação entre as variáveis originais e os componentes principais:

$$\boldsymbol{CP1} = 0.95X_1 + 0.90X_2 + 0.93X_3 + 0.96X_4 + 0.89X_5 + 0.64X_6 + 0.73X_7 + 0.79X_8 + 0.55X_9 + 0.65X_{10}$$

E&S - Engineering and Science ISSN: 2358-5390 DOI: 10.18607/ES2023121 Volume 12, Número 2

$$CP2 = 0.40X_1 - 0.23X_2 + 0.18X_3 - 0.01X_4 - 0.23X_5 - 0.65X_6 - 0.78X_7 + 0.44X_8 + 0.69X_9 + 0.46X_{10}$$

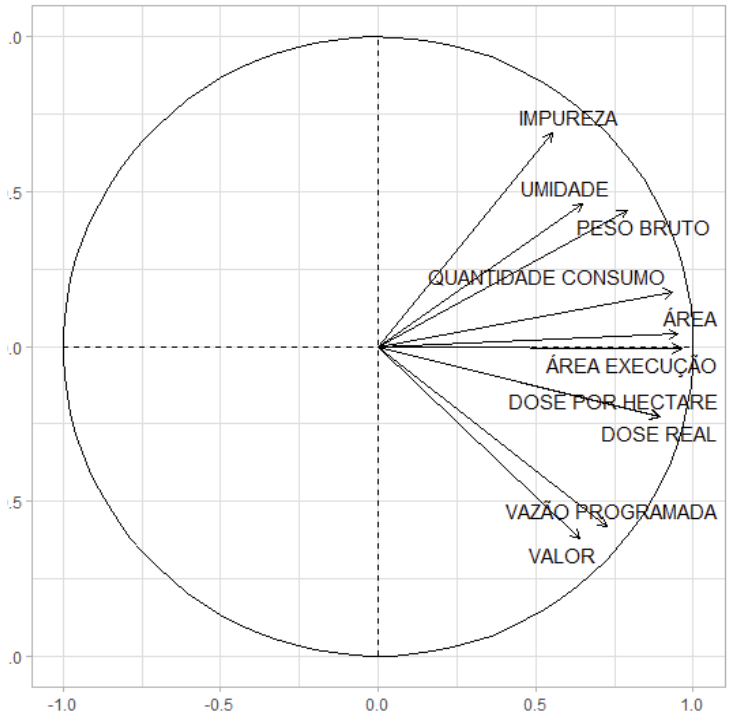
De acordo com a PC1 destacaram-se as variáveis X_1 (Área Prevista), X_2 (Dose de Fertilizante Por hectare), X_3 (Quantidade Consumo de Água (Aplicação)), X_4 (Área Execução), X_5 (Área Execução), X_6 (Dose Real) e X_{10} (Umidade), por isso esse componente foi denominado de componente de produção. Quanto ao PC2 destacaram-se as variáveis X_6 (Valor de Produção), X_7 (Vazão), X_8 (Peso Bruto), e X_9 (Impureza do Grão) e por isso podemos chamá-lo de componente de planejamento, pois todas as variáveis que o constituem são oriundas do planejamento.

As variáveis X_1 e X_4 apresentaram contribuições similares para o CP1, isto foi verificado pelas variáveis que têm vetor de maior comprimento e que foram mais próximas ao eixo CP1, mostrado na Figura 2. Existem correlações altas entre as componentes principais e seus coeficientes de ponderação de cada característica.

Com a seleção de dois componentes principais, a redução da dimensão de dez variáveis originais para dois componentes principais é bastante razoável. Portanto decidiu-se utilizar unicamente os dois primeiros componentes principais para a composição das equações 1 e 2. Não existe correlação entre as varáveis X_6 e X_9 pois forma um ângulo próximo de 90 graus, como mostrado na Figura 2.

ACP foi usada para reduzir as dimensões das variáveis originais sem perda de informação. Por definição, a correlação entre os principais componentes é zero, isto é, a variação explicada em CP1 é independente da variação explicada em CP2 e assim por diante.

Figura 2 – Biplot CP1 × CP2

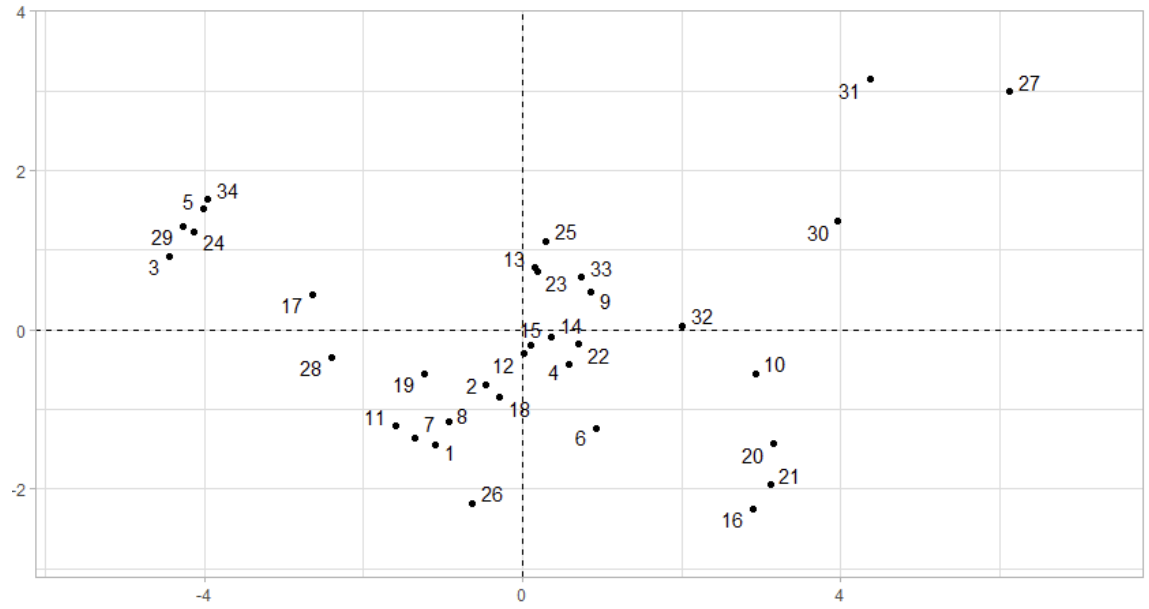


ES Engineering and Science ISSN: 2358-5390 DOI: 10.18607/ES20231215788 Volume 12, Número 3

Como os dois primeiros componentes principais explicaram 83,34% da variação total dos dados (Figura 3), o modelo se mostra adequado para utilização e aplicação uma vez que ultrapassou o recomendado (JOHNSON, 1998).

Ao analisar a Figura 3, que é o Biplot CP1 x CP2 com as variáveis de produção e planejamento (1) e (2), pode-se concluir que, de acordo com os dados de produção de soja e com a ACP, as localidades 27, 31 e 21 possuem maiores números de produção pela CP1. As localidades 3, 29, 6, 24 e 34 foram às áreas que apresentaram menores números produtividade. Ademais foi identificado que as áreas 31 e 27 foram as que tiveram melhor performance no âmbito de produção dentre as áreas analisadas, mesmo sendo áreas com baixas perspectiva nas variáveis de programação.

Figura 3 – Biplot CP1 × CP2 com os indivíduos das variáveis



CONCLUSÃO

Com base nos resultados obtidos, é possível inferir que a análise de componentes principais (ACP) revelou-se altamente eficaz. Esse procedimento permitiu a eliminação de seis variáveis com baixa variabilidade, que exibiram correlações intrínsecas ou contribuição limitada nos dados mais influentes na formação dos componentes. Em decorrência, redundâncias que poderiam comprometer análises futuras foram mitigadas.

Tal análise proporcionou a segmentação em dois componentes distintos: um referente à dados de produção (Componente de produção) e outro vinculado a dados de planejamento da área (Componente de planejamento). Esses componentes sintetizaram e simplificaram o conjunto de dados, preservando, contudo, a integridade da compreensão e interpretação. Em resultado direto, uma alocação mais precisa de recursos temporais e materiais foi efetivada, visando à otimização das práticas de cultivo e manejo do solo.

O objetivo primordial da ACP foi plenamente alcançado. Um escopo reduzido de componentes (CP1 e CP2) foi extraído contemplando uma parcela substancial da variação intrínseca nos dados primordiais (83,34%). Isso não só demonstra a exequibilidade das aplicações da análise de componentes principais em contextos do agronegócio, mas também a capacidade dessa técnica em descrição estatisticamente significativa das intricadas nuances do setor.

ES Engineering and Science ISSN: 2358-5390 DOI: 10.18607/ES20231215788 Volume 12, Número 3

REFERÊNCIAS

CAMPEÃO, P; SANCHES, A. C; MACIEL, W. R. E. Mercado Internacional de Commodities: uma análise da participação do Brasil no mercado mundial de soja entre 2008 e 2019. **Desenvolvimento em Questão**, v. 18, p.76-92, 2020. https://doi.org/10.21527/2237-6453.2020.51.76-92

FERREIRA, D. F. **Estatística Multivariada.** 2 ed. Lavras: UFLA, 2011. 233p.

FRAGA, A. B; SILVA, F. L; HONGYU, K.; SANTOS, D. D. S; MURPHY, T. W.; LOPES, F. B. Multivariate analysis to evaluate genetic groups and production traits of crossbred Holstein × Zebu cows. **Trop Anim Health Prod.**, v.2, p.533-538, 2016. https://doi:10.1007/s11250-015-0985-2.

FILHO, J. C, CAMPOS, K. C; LEMOS J. Nível tecnológico das unidades agrícolas familiares nas microrregiões do Nordeste do Brasil. **Interações (Campo Grande),** v.24, p.229-245, 2023. https://doi.org/10.20435/inter.v24i1.3771

HAIR JR., J. F; BLACK, W. C; BABIN, B. J; ANDERSON, R. E. **Análise Multivariada de Dados.** 6ª ed. Porto alegre: Bookman, 2009. 23p.

HANKE; M.S. MACHADO; S. G. S. NASCIMENTO; M. R. de AVILA; C. N. PILLON. Produção de soja sob plantio direto e convencional: análise de atributos químicos e físicos do solo. **Revista cultura agronômica**, v.31, p.64-76, 2022. https://doi.org/10.32929/2446-8355.2022v31n2p64-76

HONGYU, K. **Distribuição empírica dos autovalores associados à matriz de interação dos modelos AMMI pelo método bootstrap não-paramétrico**. 2012. Dissertação de mestrado, Escola superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.

HONGYU, K. **Comparação do GGE- biplot ponderado e AMMI-ponderado com outros modelos de interação genótipo × ambiente.** 2015. Tese de Doutorado em Estatística e Experimentação Agronômica, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis Madison: Prentice Hall International**. 6ª ed. New jersey: Pearson prentice hall, 1998.

KAISER, H. F. The varimax criterion for analytic rotation in fator analysis. **Psychometrika**, v.23, p.187-200, 1958. https://doi.org/10.1007/BF02289233