

Aplicação da análise multivariada para a produção de soja brasileira

Multivariate analysis of Brazilian soybeans production

Elton Massahiro Saito Loures ^a
Maria Cristina Neves de Oliveira ^b
Leandro Garcia Meyer ^c

Resumo: Para entender a dinâmica da produção do grão de soja, foram aplicados o método multivariado de análise de componentes principais (ACP) e a análise de Clusters (AC) para os anos de 1996 a 2017. As variáveis avaliadas foram: produção de grãos de soja brasileira, que é a variável dependente, taxa de juros, taxa de câmbio (R\$/U\$), índice de inflação, taxa de desemprego, produtividade e importação chinesa. De acordo com os anos considerados na pesquisa, foi possível concluir a importância dos dois métodos multivariados e avaliar com confiança a melhor contribuição das variáveis e suas relações com a cultura da soja, bem como identificar que as variáveis quantidade total importada de grãos de soja brasileira da China e produtividade foram as que apresentaram relação mais significativa com a produção de soja.

Palavras-chave: Produção de soja; Importação; Componentes Principais; Cluster.
Classificação JEL: Q17, C38

Abstract: In order to understand the soybean production dynamics, an analysis was carried out using multivariate methods of principal component analysis (PCA) and Cluster analysis (CA) for the years 1996 to 2017. The selected variables were: Brazilian soybean grains production, which was the dependent variable, interest rate, exchange rate (R\$/U\$), inflation index, unemployment rate, yield and Chinese imports. It was possible to conclude that two applied multivariate methods allowed to confidently evaluate the best contribution of the variables and their relationships according to the years considered in the research with the soybean culture, as well as identify that variables total quantity of Brazilian soybean imported from China and productivity were the ones that showed the most significant relationship with soybean production.

Keywords: Soybeans production; Import; Principal Components; Clusters.
JEL Classification: Q17, C38

^a Discente de Ciências Econômicas-UEL. E-mail: masssahirosaito@gmail.com.

^b Pesquisadora da Embrapa Soja. E-mail: mcristina.oliveira@embrapa.br.

^c Professor colaborador-UEL. E-mail: leandro.meyer@uel.br.

1. Introdução

A soja (*Glycine max* (L.) Merrill) é uma das mais importantes culturas da economia mundial. Seus grãos são muito utilizados pela agroindústria, indústria química e de alimentos e recentemente como fonte alternativa de biocombustível (COSTA; ROSSI, 2000). Nas últimas décadas a produção de soja está entre as atividades econômicas que apresentaram crescimentos mais expressivos, o que pode ser atribuído a fatores sobretudo de natureza mercadológica e tecnológica (HIRAKURI *et al.*, 2019). Atualmente, a soja é o principal produto da agricultura brasileira. Ela fortalece a posição do país como um dos principais produtos e exportadores do comércio agrícola mundial, proporcionando ao Brasil a capacidade de influenciar o preço do mercado mundial de *commodities* agrícolas (HIRAKURI, LAZZAROTTO, 2014). Segundo a Conab (2019), a área cultivada desta *commodity* no Brasil chegou a 36,437 milhões de hectares, o que, considerando a produtividade média dos últimos três anos de 3.352kg/ha corresponde a uma produção de 122,13 milhões de toneladas, com exportação estimada em 72 milhões de toneladas.

Neste contexto, destaca-se o papel desempenhado pela China em termos de comércio mundial da soja e também no que se refere às relações comerciais com o Brasil. Dados da Comtrade Group apresentados por Ramos *et al.* (2020), mostram que a participação das importações chinesas nas importações totais de soja no mundo variou entre 38 e 40% entre 2012 e 2016. Este número torna-se ainda mais expressivo ao se levar em conta que a Holanda, segundo país com maior participação na importação total de soja, apresentou participação de 4% ao longo do mesmo período.

Destaca-se que o Brasil divide com os Estados Unidos da América (EUA) a maior participação nas exportações mundiais totais entre 2012 e 2016, com ambos países apresentando participação entre 27% e 32% em todos os anos, conforme dados apresentados por Ramos *et al.* (2020). De acordo com dados da Estatísticas de Comércio Exterior (COMEX, 2020), em 2019 o Brasil exportou aproximadamente 26 USD bilhões de soja em grão para o mundo todo, sendo que desse montante 20,5 USD tiveram a China como destino, o que corresponde a uma participação de aproximadamente 78% das exportações totais de soja do Brasil. Levando-se em conta que, de acordo com a mesma fonte, o Brasil exportou um total de 225 USD bilhões em 2019, somente a exportação de soja para a China representou aproximadamente 11% das exportações totais brasileiras no ano.

Por conta de tamanha relevância da soja na economia brasileira, há grande quantidade de estudos a respeito das safras de cada ano, bem como um abundante número de pesquisas imersivas do cenário empresarial na literatura acadêmica e nos noticiários da área econômica. Porém, poucas publicações têm discutido a relação entre a produção brasileira de soja e o conjunto de variáveis de ordem econômica, monetária e a relação comercial de outros países, com destaque para a China.

Visando oferecer maiores informações a este respeito, o objetivo deste estudo foi verificar a importância de aplicar métodos multivariados e analisar a produção de soja no

Brasil com respeito às variáveis de ordem econômica e monetária e a quantidade total de importação de grãos de soja brasileira pela China, referente aos anos de 1996 a 2017. Para tanto, foi aplicada a análise de componentes principais (ACP) que permite reduzir a dimensão do número de variáveis com o mesmo grau de confiabilidade do conjunto de variáveis originais. Também foi aplicada a Análise de Cluster (AC) no intuito de dividir em grupos similares os elementos da amostra.

A ACP foi desenvolvida com base nos artigos de Hotelling (1933) e de Pearson (1901), sendo uma técnica bastante utilizada em diversas áreas do conhecimento, como a biologia, a agronomia, a zootécnica, a ecologia, a engenharia florestal, a medicina, a economia, entre outras áreas. Sugere-se o seu uso quando o volume de dados e ou variáveis é grande, possibilitando reduzir a dimensão da matriz de dados que compõem o conjunto de variáveis resposta, resultando em apenas poucos componentes.

Já a AC é uma técnica que tem como propósito dividir em grupos os elementos da amostra ou da população, de modo que as informações pertencentes a um mesmo grupo sejam similares entre si com respeito às características medidas (MINGOTI, 2007; MANLY, 1994). A AC pode ser aplicada em diversas áreas, como por exemplo na psicologia para a classificação de pessoas de acordo com seus perfis de personalidade, posicionamento de produtos numa pesquisa de mercado, entre outras possibilidades. O principal benefício de tal metodologia consiste na redução da dimensionalidade da amostra, auxiliando em problemas de classificação. Por meio do método hierárquico de Ward (1963), que é um método de agrupamento fundamentado na mudança de variação dentro dos grupos de formação, e entre eles, foi elaborada a análise de agrupamentos. Segundo Romesburg (1984), este método possui visualização de grupos bem definidos no dendograma, o que o torna um método hierárquico preferido na presente pesquisa.

Assim, este artigo está organizado em cinco seções além da presente introdução. Na seção seguinte é apresentado a revisão teórica. Na terceira e na quarta seção apresentam-se o método empregado e os resultados e discussão. Por fim, na quinta seção são destacadas as conclusões.

2. Revisão teórica

Segundo Oliveira (2018), há poucas publicações que discutem especificamente a relação entre a produção agrícola do Brasil e o conjunto de variáveis de ordem monetária, econômica, e de renda. Kich, Coronel e Vieira (2012) verificaram no período entre 1997 a 2009 as influências das variáveis macroeconômicas no saldo da balança comercial do agronegócio brasileiro e concluíram que as variáveis renda e taxa de câmbio foram as mais significativas do saldo da balança. Oliveira (2018) concluiu em sua pesquisa que a variação da produção brasileira de soja é quase totalmente explicada pelas alterações do PIB, da taxa de câmbio e da taxa SELIC. Conforme a pesquisa desenvolvida por Najberg e Pereira (2004), a grande contribuição do agronegócio possui para a geração de empregos no país. Especialmente em relação ao complexo de soja brasileira, Roessing e

Lazzarotto (2004) apresentam que esse complexo é responsável por gerar em torno de cinco milhões de postos de trabalho.

Entre 1995 e 2005, o comércio brasileiro dobrou com o mundo e com a China aumentou cinco vezes, refletindo um forte crescimento comercial entre os dois países, tanto das importações quanto das exportações (Costas, 2015). Segundo a USDA-FAS (2019), o Brasil (35,44%), EUA (30,15%) e Argentina (15,27%) seriam responsáveis por 80,87% de toda a produção mundial de soja em grão, ao passo que a China responderia por 57,68% de todas as importações mundiais. Freitas (2013) analisou os principais mercados consumidores das exportações de soja brasileira de 1995 a 2010 e foi possível verificar elevado grau de concentração das exportações para a China – em 2010, a China importou por mais de 60% de todo o valor exportado do Brasil. Conforme Saito (2019), foi possível verificar que para a produção de grãos de soja brasileira, é importante a relação comercial externa, principalmente com a China e sua demanda pelo produto.

3. Metodologia

3.1. Análise de Componentes Principais (ACP) e Biplot

Para aplicar esta metodologia, os componentes principais devem ser definidos como combinações lineares de p variáveis correlacionadas e, cada combinação linear é não-correlacionada com as demais variáveis (HOFFMANN, 2011). Essas combinações lineares são em número menor do que o número de variáveis originais presentes na pesquisa (REIS, 2001, FERREIRA; 2008). Essa metodologia agrupa indivíduos conforme sua variância, ou seja, de acordo com seu comportamento dentro da população. Sendo assim, é conveniente padronizar cada variável, permitindo que todas as variáveis tenham o mesmo peso (HONGYU *et al.*, 2015).

Seja $X = (X_1, X_2, \dots, X_p)'$ um vetor aleatório com vetor de médias $\mu = (\mu_1, \mu_2, \dots, \mu_p)'$ e matriz de correlações $\Sigma_{p \times p}$. Sejam os autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ e seus respectivos autovetores e_1, e_2, \dots, e_p normalizados. A matriz ortogonal da dimensão $p \times p$, constituída dos autovetores normalizados da matriz $\Sigma_{p \times p}$ pode ser representada por:

$$O_{p \times p} = (e_{11} \ e_{21} \ \dots \ e_{p1} \ e_{12} \ e_{22} \ \dots \ e_{p2} \ \dots \ e_{1p} \ e_{2p} \ \dots \ e_{pp}) = [e_1 \ e_2 \ \dots \ e_p] \quad (1)$$

O vetor Y é composto de p combinações lineares das variáveis aleatórias do vetor X , com médias igual a $O' \mu$ e matriz de correlações $\Lambda_{p \times p}$, que é uma matriz diagonal, com elementos iguais a $a_{ii} = \lambda_i, i = 1, 2, \dots, p$, isto é,

$$\Lambda_{p \times p} = [\lambda_1 \ 0 \ \lambda_2 \ 0 \ \lambda_p] \quad (2)$$

As variáveis aleatórias do vetor Y não são correlacionadas entre si. O que leva à ideia do uso de combinações lineares em Y como forma alternativa de representar a estrutura de covariâncias do vetor X . As componentes principais serão determinadas a partir da matriz de covariâncias das variáveis originais padronizadas (MINGOTI, 2007), onde $Z = (Z_1, Z_2, \dots, Z_p)'$ é um vetor aleatório padronizado com os respectivos autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ da matriz de correlação $P_{p \times p}$ e seus respectivos autovetores e_1, e_2, \dots, e_p normalizados, ou seja, as componentes principais são determinadas a partir da matriz de correlações.

Para o método ACP, foram utilizados os softwares GGE biplot (YAN; KANG, 2003) e Statistica (STATSOFT, 2007). O gráfico biplot permite verificar a contribuição dos diferentes anos de pesquisa proposto e avaliar as variáveis resposta. Esta metodologia, apresentada por Yan Rajcan (2002) e Yan (2014), permite a decomposição de valor singular (DVS) dos dois primeiros componentes principais, o que significa que uma matriz é decomposta em duas partes: autovetores de valores singulares de colunas, e autovetores de linhas, formando uma matriz diagonal. Inicialmente este método foi desenvolvido por Gabriel (1971) e pode interpretar múltiplas variáveis em função de diferentes tratamentos no mesmo gráfico (AKINWALE *et al.*, 2014). O modelo para avaliar estes resultados e obter o gráfico biplot poligonal é baseado na Decomposição do Valor Singular (DVS) dos componentes principais, e a expressão para o enfoque multivariado está na Equação 3:

$$\frac{T_{ij} - T_j}{S_j} = \lambda_1 \varphi_{1i} \tau_{1j} + \lambda_2 \varphi_{2i} \tau_{2j} + \varepsilon_{ij} \quad (3)$$

em que: T_{ij} é o efeito do ano i na variável j ; T_j é o valor médio do ano agrícola na variável j ; S_j , o desvio padrão da interação do ano sob pesquisa (tratamento) i e a variável j nas colunas; φ_{1i} e φ_{2i} , os escores dos componentes CP1 e CP2, respectivamente, para o ano de pesquisa i ; τ_{1i} e τ_{2j} , os escores de CP1 e CP2, respectivamente para a variável j ; e ε_{ij} , o efeito residual associado ao modelo com a interação do ano e as variáveis nas colunas j .

Antes de qualquer procedimento multivariado, deve-se avaliar a existência de coeficiente de correlação entre as variáveis igual ou superior a $r = 0,30$, pois é da referida matriz que são extraídos os autovalores e os autovetores e , conforme Hair *et al.* (2014), valores inferiores a este valor podem representar um preliminar indício de serem inapropriados. As novas variáveis não correlacionadas, obtidas de combinações lineares das variáveis originais, são apresentadas em ordem decrescente relativas ao seu poder de explicação. A ACP é bem sucedida se existem fortes intercorrelações nos dados originais e se estes estão padronizados para uma mesma escala relativa, de forma que os eixos da ACP não sejam dominados por uma ou mais variáveis que apresentam as maiores variâncias. Caso contrário, não se consegue capturar uma série de novas variáveis, pois a dimensão da matriz é reduzida com perda de informações e, com isso, não se consegue explicar a maior parte da variação nos dados (GOTELLI; ELLISSON, 2011; HOFFMANN, 2011).

3.2. Análise de Cluster (AC) pelo método de Ward

O método proposto por Ward (1963) é um método de agrupamento fundamentado na mudança de variação dentro dos grupos de formação e entre cada passo do processo de agrupamento. Para cada possibilidade de agrupamento é calculado a soma de quadrados total dentro dos grupos, em seguida, une-se os grupos que resultam no menor valor dessa soma de quadrados. Portanto, como este método busca a partição que minimiza a variabilidade dentro dos grupos elaborados, é conhecido como “mínima variância” (SILVA, 2013). Conforme Barroso e Artes (2003), o método de Ward é um método atraente por gerar grupos que possuem uma alta homogeneidade interna. Para aplicar este método foi utilizado o software Statistica (STATSOFT, 2007).

Inicialmente cada elemento é considerado como um único conglomerado e em cada passo do algoritmo de agrupamento é calculada a soma de quadrados dentro de cada conglomerado. O agrupamento é feito a partir das somas de quadrados dos desvios entre acessos ou a partir do quadrado da distância Euclidiana.

$$SS_i = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)' (X_{ij} - \bar{X}_i) \quad (4)$$

em que n_i é o número de elementos no conglomerado C_i quando se está no passo k do processo de agrupamento; X_i , o vetor de observações do j -ésimo elemento amostral que pertence ao i -ésimo conglomerado; \bar{X}_i , o centroide do conglomerado; e SS_i , a soma de quadrados correspondente do conglomerado C_i . Ao passo k , a soma de quadrados total dentro dos grupos é representada como:

$$SSR = \sum_{i=1}^{g_k} SS_i \quad (5)$$

onde g_k é o valor de grupos existentes quando se está no passo k .

Portanto a distância entre os conglomerados C_l e C_i é definida pela soma dos quadrados entre os clusters C_l e C_i :

$$d(C_l, C_i) = \left[\frac{n_l n_i}{n_l + n_i} \right] (\bar{X}_l - \bar{X}_i)' (\bar{X}_l - \bar{X}_i) \quad (6)$$

em que a cada passo do algoritmo de agrupamento, os dois conglomerados que minimizam a distância são combinados.

O método de Ward leva em consideração a diferença dos tamanhos dos conglomerados que são comparados. As comparações que possuem tamanhos diferentes sofrem uma penalização representada pelo fator de ponderação $\frac{n_l n_i}{n_l + n_i}$. Quanto maiores forem os valores de n_l e n_i e a discrepância entre eles, maior será o fator de ponderação

e, portanto, maior a distância entre os centroides dos conglomerados comparados (MINGOTI, 2007).

3.3. Variáveis utilizadas

As variáveis resposta avaliadas foram a produção de soja (PRODSOJA), que é a variável dependente, isto é, a variável cuja compreensão do comportamento constitui o objetivo da presente pesquisa, e as demais variáveis independentes, que foram a taxa SELIC (TXSELIC), a taxa de câmbio (TXCAMBIO), a taxa IGPM (TXIGPM), a taxa de desemprego (TXDESEMP), a quantidade total de importação de grãos de soja brasileira da China (IMPCHINA) e a produtividade (PRODUTIVIDADE), conforme consta no Quadro 1. Os dados deste trabalho, são referentes ao período de 1996 a 2017 de valores acumulados no ano. Todas as variáveis resposta foram padronizadas a fim de equilibrar os valores de variância e colocar os dados na mesma escala de medida.

Quadro 1. Caracterização das variáveis.

Taxa básica de juros (TXSELIC)	Taxa média ajustada (TMA) dos financiamentos diários, acumulado no ano, apurados no Sistema Especial de Liquidação e de Custódia (SELIC) e lastreadas em títulos federais (BANCO CENTRAL DO BRASIL, 2020).
Taxa de câmbio (TXCAMBIO)	Preço da moeda estrangeira medido na moeda nacional (R\$/US\$) (BANCO CENTRAL DO BRASIL, 2020).
Taxa de desemprego (TXDESEMP)	Porcentagem de pessoas na força de trabalho que estão desocupadas (IBGE, 2020).
Índice Geral de Preços do Mercado (TXIGPM)	Medido pela Fundação Getúlio Vargas (FGV), indicador que registra a inflação de preços (BANCO CENTRAL DO BRASIL, 2020).
Importação de grãos de soja da China (IMPCHINA)	Quantidade total de importação de grãos de soja brasileira da China medida em milhares de toneladas (USDA-FAS, 2019).
Produtividade (PRODUTIVIDADE)	Razão da produção total de grãos de soja brasileira e a área plantada de grãos de soja medida em hectares (IBGE, 2020).
Produção de Soja (PRODSOJA)	Quantidade total da produção de grãos de soja no país em milhares de toneladas (USDA-FAS, 2019).

Fonte: Elaboração própria.

4. Resultados e discussão

4.1. Coeficientes de Correlação de Pearson para o período estudado

Para a realização e aplicação do método multivariado, se faz necessário observar a existência de um determinado padrão de correlação entre as variáveis e que pode ser avaliado pelo método de Pearson. Como a matriz de Correlação de Pearson entre as variáveis estudadas apresentou um padrão de correlação superior a 0,30, foi possível ao

estudo a aplicação dos métodos ACP e AC. O Quadro 2 apresenta o coeficiente de Correlação de Pearson para as sete variáveis pesquisadas, o que permite constatar que as variáveis TXDESEMP e TXIGPM apresentaram os menores valores do referido coeficiente em relação as demais. Dos 21 coeficientes possíveis de correlação, apenas 10 foram significativos pelo teste $t \leq 0,05$.

Quadro 2. Coeficientes de Correlação de Pearson entre as variáveis econômicas da soja no Brasil no período de 1996 a 2017.

Variáveis	PRODS OJA	IMPCH INA	TXCAM BIO	TXSE LIC	TXDESE MP	PRODUTIV DADE	TXIG PM
PRODSOJA	1,00	0,98*	0,62*	-0,69*	0,13	0,85*	-0,35
IMPCHINA	0,98	1,00	0,55*	-0,70*	0,07	0,82*	-0,38
TXCAMBIO	0,62	0,55	1,00	-0,40	0,53*	0,45*	0,10
TXSELIC	-0,69	-0,70	-0,40	1,00	0,03	-0,63*	0,29
TXDESEMP	0,13	0,07	0,53	0,03	1,00	0,08	0,01
PRODUTIVI DADE	0,85	0,82	0,45	-0,63	0,08	1,00	-0,20
TXIGPM	-0,35	-0,38	0,10	0,29	0,01	-0,20	1,00

Fonte: Resultados da pesquisa.

* Valores de coeficientes de correlação significativos pelo teste $t \leq 0,05$.

Nota: Quantidade total produzido de grãos de soja brasileira (PRODSOJA), Quantidade total de importação de grãos de soja brasileira pela China (IMPCHINA), Taxa de Câmbio (TXCAMBIO), Taxa básica de juros SELIC (TXSELIC), Taxa de Desemprego (TXDESEMP), Produtividade calculada pela razão da quantidade total da produção de soja e área colhida em hectares (PRODUTIVIDADE), Índice Geral de Preços do Mercado (TXIGPM).

4.2. Análise de Componentes Principais (ACP)

Nesse estudo foram indicados dois componentes, observando a proposta de aceitar somente os autovalores iguais ou superiores à unidade (KAISER, 1960). Os autovalores somaram 75,42% da variação total (CP1= 55,29% e CP2= 20,14%). No Quadro 3 foram apresentados os autovalores, bem como as percentagens para os dois primeiros componentes.

Quadro 3. Autovalores obtidos da matriz de correlação na pesquisa de análise da produção de grãos de soja com variáveis de ordem econômica monetária e a importação chinesa de grãos.

Componentes	Autovalores	% da variância total	Variância acumulada	% da variância acumulada
CP1	3,870	55,287	3,870	55,287
CP2	1,410	20,138	5,280	75,425

Fonte: Resultados da pesquisa.

Os autovetores definem a direção da máxima variabilidade¹ e por meio deles é possível caracterizar os componentes principais. Foram observados no Quadro 4, em ordem decrescente, que os maiores valores absolutos de escores ocorreram no primeiro componente (CP1) para as variáveis PRODSOJA (-0,496), IMPCHINA (-0,487), PRODUTIVIDADE (-0,443), TXSELIC (0,401) e TXCAMBIO (-0,340). Ao segundo componente (CP2), em ordem decrescente as variáveis que mais relacionam foram: TXDESEMP (0,680), TXCAMBIO (0,539) e TXIGPM (0,442). Os demais valores não foram significantes (Quadro 4).

Quadro 4. Valores calculados para os autovetores baseando-se na matriz de correlação, nos dois primeiros componentes principais com as variáveis sob pesquisa, no período de 1996 a 2017.

Variáveis	CP1	CP2
PRODSOJA	-0,496	-0,037
IMPCHINA	-0,486	-0,107
TXCAMBIO	-0,340	0,539
TXSELIC	0,401	0,185
TXDESEMP	-0,105	0,680
PRODUTIVIDADE	-0,443	-0,065
TXIGPM	0,183	0,442

Fonte: Resultados da pesquisa.

Nota: Quantidade total produzido de grãos de soja brasileira (PRODSOJA), Quantidade total de importação de grãos de soja brasileira pela China (IMPCHINA), Taxa de Câmbio (TXCAMBIO), Taxa básica de juros SELIC (TXSELIC), Taxa de Desemprego (TXDESEMP), Produtividade calculada pela razão da quantidade total da produção de soja e área colhida em hectares (PRODUTIVIDADE), Índice Geral de Preços do Mercado (TXIGPM).

Sendo assim, pode-se escrever os dois primeiros componentes \hat{Y}_1 e \hat{Y}_2 conforme as expressões abaixo:

$$\hat{Y}_1 = -0,496(\text{PRODSOJA}) + 0,401(\text{TXSELIC}) - 0,340(\text{TXCAMBIO}) - 0,183(\text{TXIGPM}) + 0,105(\text{TXDESEMP}) + 0,486(\text{IMPCHINA}) + 0,443(\text{PRODUTIVIDADE})$$

$$\hat{Y}_2 = -0,037(\text{PRODSOJA}) + 0,185(\text{TXSELIC}) + 0,539(\text{TXCAMBIO}) + 0,442(\text{TXIGPM}) + 0,680(\text{TXDESEMP}) - 0,107(\text{IMPCHINA}) - 0,065(\text{PRODUTIVIDADE})$$

¹ Dependendo do *software* utilizado e do determinado modelo de formulação de componentes principais, pode resultar em sinais diferentes entre os resultados, esta diferença se trata de uma reflexão em relação ao eixo. Uma rotação em seu espaço vetorial n-dimensional em torno da origem, poderá ocasionar uma "rotação" em torno do eixo (HERIG, 1979).

Por meio da correlação entre uma variável e uma componente principal, baseados na matriz de correlação (Quadro 5), o primeiro componente (CP1) apresentou altos coeficientes de correlação entre PRODSOJA (-0,98), IMPCHINA (-0,96), PRODUTIVIDADE (-0,87), TXSELIC (0,79) e TXCAMBIO (-0,66), permitindo verificar uma relação de algumas variáveis propostas de ordem econômica e a quantidade total de importação de grãos de soja brasileira pela China com a produção e produtividade de grãos de soja brasileira. Ao caso das variáveis TXDESEMP (-0,20) e TXIGPM (0,36) resultaram em valores baixos. A segunda componente com coeficientes de correlação importantes foram o conjunto de variáveis de ordem econômica e monetária padronizadas TXDESEMP (0,81), TXCAMBIO (0,64), TXIGPM (0,52), com exceção da variável taxa básica de juros TXSELIC (0,185).

Quadro 5. Coeficientes de correlação entre os componentes principais e as variáveis originais (Loadings), baseados na matriz de correlação, no período de 1996 a 2017.

Variáveis	CP1	CP2
PRODSOJA	-0,98	-0,04
IMPCHINA	-0,96	-0,13
TXCAMBIO	-0,66	0,64
TXSELIC	0,79	0,22
TXDESEMP	-0,20	0,81
PRODUTIVIDADE	-0,87	-0,08
TXIGPM	0,36	0,52

Fonte: Resultados da pesquisa.

Nota: Quantidade total produzido de grãos de soja brasileira (PRODSOJA), Quantidade total de importação de grãos de soja brasileira pela China (IMPCHINA), Taxa de Câmbio (TXCAMBIO), Taxa básica de juros SELIC (TXSELIC), Taxa de Desemprego (TXDESEMP), Produtividade calculada pela razão da quantidade total da produção de soja e área colhida em hectares (PRODUTIVIDADE), Índice Geral de Preços do Mercado (TXIGPM).

As variáveis que mais contribuíram para atender o objetivo no CP1 apresentaram em ordem decrescente os seguintes escores: PRODSOJA (0,246), IMPCHINA (0,237) e PRODUTIVIDADE (0,196), TXSELIC (0,161), TXCAMBIO (0,116). Os demais valores foram inexpressivos. No CP2 os maiores escores percentuais foram: TXDESEMP (0,463), a TXCAMBIO (0,290) e TXIGPM (0,195). As demais variáveis não apresentaram contribuição importante (Quadro 6).

Quadro 6. Contribuições das variáveis na matriz de correlação, nos dois primeiros componentes principais com as variáveis sob pesquisa, no período de 1996 a 2017.

Variáveis	CP1	CP2
PRODSOJA	0,246	0,001
IMPCHINA	0,237	0,012
TXCAMBIO	0,116	0,290
TXSELIC	0,161	0,034
TXDESEMP	0,011	0,463
PRODUTIVIDADE	0,196	0,004
TXIGPM	0,034	0,195

Fonte: Resultados da pesquisa.

Nota: Quantidade total produzido de grãos de soja brasileira (PRODSOJA), Quantidade total de importação de grãos de soja brasileira pela China (IMPCHINA), Taxa de Câmbio (TXCAMBIO), Taxa básica de juros SELIC (TXSELIC), Taxa de Desemprego (TXDESEMP), Produtividade calculada pela razão da quantidade total da produção de soja e área colhida em hectares (PRODUTIVIDADE), Índice Geral de Preços do Mercado (TXIGPM).

Foram observados que alguns anos tiveram destaque e podem ser elencados em ordem decrescente (Quadro 7), no CP1: 2017 (25,302), 1997 (15,301), 1996 (11,278), 2016 (9,719), 2015 (8,043), 1998 (7,456) e 1999 (5,411). No componente principal 2, os anos que mais contribuíram foram: 2002 (18,854), 2016 (10,839), 2011 (10,111), 1999 (7,269), 2014 (7,090), 2013 (6,088), 2003 (5,436), 2017 (5,044), 2012 (4,867), 2004 (4,677) e 2009 (4,583).

Quadro 7. Contribuição dos tratamentos (anos) baseando-se na matriz de correlação, nos dois primeiros componentes principais com as variáveis sob pesquisa, no período de 1996 a 2017.

Ano	CP1	CP2
1996	11,278	3,581
1997	15,307	0,383
1998	7,456	0,656
1999	5,411	7,269
2000	2,656	0,881
2001	0,837	2,164
2002	2,605	18,854
2003	0,079	5,436
2004	0,831	4,677
2005	0,532	0,132
2006	0,217	0,446

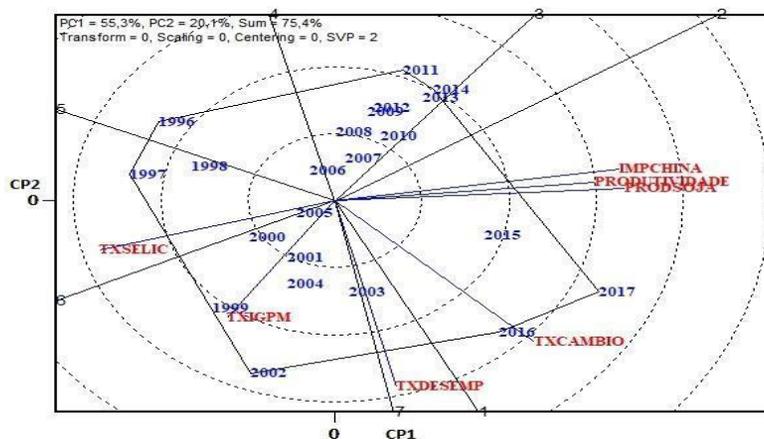
2007	0,043	1,005
2008	0,001	2,627
2009	0,401	4,582
2010	0,752	2,348
2011	1,707	10,111
2012	0,566	4,867
2013	2,757	6,088
2014	3,498	7,090
2015	8,043	0,917
2016	9,719	10,839
2017	25,302	5,044

Fonte: Resultados da pesquisa.

No biplot poligonal as linhas perpendiculares são linhas de igualdade entre os anos adjacentes no polígono, o que facilita a comparação visual deles. As linhas de igualdade dividem o biplot poligonal em setores, e o tratamento (anos) vencedor para cada setor é aquele localizado no vértice (YAN; TINKER, 2006). No biplot o comprimento do vetor é medido em relação à distância da origem, sendo a diferença do valor da variável calculada ao considerar as médias das variáveis, ou seja, a partir do comprimento do vetor, é possível visualizar interações específicas entre as variáveis utilizadas e dos tratamentos e quanto maior o comprimento do vetor, maior será a contribuição do tratamento e da variável (HONGYU *et al.*, 2015).

Considerando que os dois componentes representaram 75,4% da variação total, o biplot avalia os anos de 1996-2017 e as variáveis utilizadas na análise. Foi observado que os anos que foram mais responsivos, com os maiores valores para as variáveis propostas foram os anos de 1996, 1997, 2002, 2011, 2013, 2014, 2016 e 2017, pois estão localizadas nos vértices do biplot (Figura 1).

As linhas que se propagam no gráfico dividem o biplot em setores e, para cada setor, há uma variável de ordem econômica, monetária ou relacionada ao volume de grãos de soja no vértice, que apresenta melhor desempenho. No biplot as sete variáveis estão divididas em três grupos, as quais se enquadram entre sete setores (Figura 1).

Figura 1: Biplot poligonal aos anos de 1996-2017

Fonte: Resultados da pesquisa.

Nota: Componentes Principais entre as variáveis econômicas (quantidade total da produção de soja em milhares de toneladas de grãos -PRODSOJA, Taxa Selic -TXSELIC, Taxa de Câmbio - TXCAMBIO, Taxa IGPM - TXIGPM, Taxa de Desemprego - TXDESEMP, quantidade total de importação de grãos de soja brasileira pela China em milhares de toneladas – IMPCHINA e Produtividade – PRODUTIVIDADE) padronizadas e os diferentes anos de avaliação (1996 a 2017). CP1= 55,3% e CP2= 20,1% e CP Total= 75,4%.

O primeiro grupo do setor 1-2 (linhas de radiação) é composto pelas variáveis relacionadas a TXCAMBIO e fortemente associado ao ano 2016, que está no vértice. Segundo De Lima Filho (2016), o mercado de soja em 2016 foi marcado pela influência da supersafra norte-americana e pela taxa de câmbio, ano em que a queda do dólar em relação ao real aumentou a pressão de baixa no mercado brasileiro. Ainda neste setor e próximo do tratamento no vértice, é possível observar que o ano de 2017 próximo das variáveis produção de grãos de soja, IMPCHINA, PRODUTIVIDADE, PRODSOJA e TXCAMBIO. Segundo Lemos (2017), em 2017 houve aumento da oferta impulsionado pelo significativo ganho de produtividade devido à adoção de novas variedades de sementes, à aplicação de fertilizantes, pesticidas e práticas de manejo, além da melhoria dos padrões de vida em economias emergentes, em particular a China, intensificando a demanda de proteínas cujo sistemas de produção são dependentes da soja. As variáveis IMPCHINA, PRODUTIVIDADE e PRODSOJA se encontram próximo do quarto círculo concêntrico. Podemos verificar no primeiro setor uma correlação da importação chinesa de grãos de soja, a taxa de câmbio (TXCAMBIO), a produção de grãos de soja e a produtividade do Brasil (Figura 1).

Conforme Yan e Tinker (2006), tratamentos localizados no lado esquerdo do biplot, são questionáveis, a não ser que exista uma justificativa importante, ou sua contribuição será muito boa ou muito pobre. Desse modo os anos de 1996, 1997 e 2002 são questionáveis e demandam um estudo mais aprofundado de sua contribuição (Figura 1). Esses autores

afirmam ainda, que o comprimento do vetor, que é a distância entre um tratamento e a origem do biplot, mede a diferença do tratamento em relação ao tratamento “médio”. Portanto, tratamentos ou variáveis com os vetores mais longos são os melhores ou os mais pobres tratamentos. Embora esses vetores estejam localizados no vértice do polígono, nem sempre são a melhor resposta. Se estiverem localizados no lado esquerdo do biplot, esses tratamentos mostram os piores valores e deve-se tomar cuidado para não ter uma interpretação equivocada, como no caso da TXSELIC e TXIGPM (Figura 1).

Os resultados apresentados corroboram com as evidências disponíveis na literatura de Souza (2018) e Souza *et al.* (2019), que por meio de Modelos de Autorregressão Vetorial com Correção de Erros (VEC) destacam, no período de janeiro de 2005 a dezembro de 2015, uma expressiva influência positiva do câmbio e do preço de exportação sobre as exportações de soja, ao passo que as variações no preço doméstico do grão e na renda causam impactos negativos sobre as exportações; verificando também que a renda é a variável que apresenta maior influência ao processamento de soja no Brasil. Com o mesmo método, entre 1980 e 2001, Figueiredo *et al.* (2004) verificou que as exportações reagem às variações no preço externo, preço interno e a renda interna com elasticidades altas e significativas. Da mesma forma, no mesmo trabalho a taxa de câmbio também apresentou, com elasticidade positiva e significativa, que as recorrentes desvalorizações no período de câmbio flutuante incentivaram as exportações de soja em grãos.

Nas linhas de radiação de 2-5, encontram-se nos vértices os anos 2013, 2014, 2011 e 1996, mas esses tratamentos não estão associados a qualquer variável resposta estudada (Figura 1). No setor 5-6 está a variável de ordem econômica TXSELIC, localizada no terceiro círculo concêntrico, que está presente e associada aos anos de 1997 no vértice e um pouco distanciado no segundo círculo concêntrico o ano de 1998. Por meio do biplot é possível verificar que a taxa SELIC padronizada com a relação da cultura de soja brasileira e o mercado externo em valores padronizados se encontram em sentidos opostos, o que confirma sua correlação negativa com as variáveis (PRODUTIVIDADE, IMPCHINA E PRODSOJA). Nas linhas de radiação ou no setor 6-7 a TXIGPM está associada ao ano de 1999 e o ano de 2002 que se encontra no vértice. Segundo Quadros (2003), os anos de 1999 e 2002 tiveram algumas semelhanças embora distribuídos no tempo de forma diferente, como por exemplo o choque cambial de mesma magnitude. Conforme Menezes (2013), a taxa IGPM possui reação muito rápida e intensa a estes choques cambiais.

O terceiro grupo, composto pela variável enquadrada nas linhas de radiação 7-1 está a TXDESEMP associada ao ano de 2003 que, está no segundo círculo concêntrico. Por meio do biplot pode-se observar a correlação das taxas cambiais e a taxa de desemprego no Brasil (Figura 1).

A metodologia de ACP foi eficiente na redução do número de variáveis, com o mesmo grau de confiabilidade do conjunto de variáveis, tornando possível avaliar com confiança a melhor contribuição das variáveis de ordem econômica, monetária e a quantidade de importação total de grãos de soja brasileira pela China e suas relações de acordo com os anos considerados na pesquisa com a cultura da soja, permitindo identificar que as variáveis IMPCHINA e PRODUTIVIDADE foram as mais representativas no

estudo. A análise corrobora com Souza e Bittencourt (2019) que avaliam o período de 1997 a 2016 e destacam o efeito do comércio internacional como o principal responsável pela expansão da cultura da soja no Brasil, com destaque para a importação chinesa. Além disso foi possível verificar que a análise biplot permite uma análise significativa e eficiente para agrupamentos, auxiliando o pesquisador por meio da análise visual.

4.3. Análise de *Cluster* (AC)

O Quadro 8 apresenta a primeira matriz de similaridade ou distância euclidiana, conforme ocorrem os agrupamentos; novas distâncias são calculadas com base nessa matriz; ela indica a distância entre as variáveis propostas na presente pesquisa. Como é uma matriz simétrica, pode-se observar somente uma das diagonais. A menor distância euclidiana entre as variáveis pesquisadas se encontra em PRODSOJA-IMPCHINA (0,82), PRODSOJA-PRODUTIVIDADE (2,50) e IMPCHINA-PRODUTIVIDADE (2,74) que foram as primeiras a serem agrupadas dentre todas as outras variáveis.

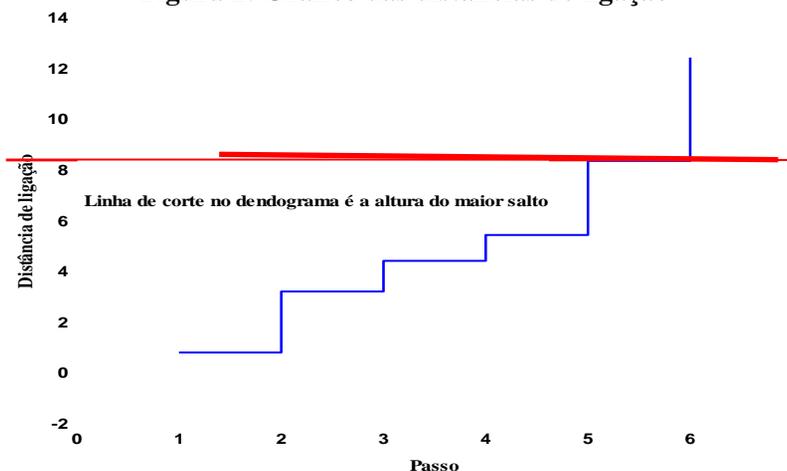
Quadro 8. Distância euclidiana entre as variáveis pesquisadas no período de 1996 a 2017

Variáveis	PRODS OJA	IMPCH INA	TXCAM BIO	TXSE LIC	TXDESE MP	PRODUTIV DADE	TXIG PM
PRODSOJA	0,00	0,82	3,97	8,42	6,03	2,50	7,53
IMPCHINA	0,82	0,00	4,34	8,45	6,26	2,74	7,60
TXCAMBIO	3,97	4,34	0,00	7,67	4,43	4,81	6,14
TXSELIC	8,42	8,45	7,67	0,00	6,38	8,28	5,45
TXDESEMP	6,03	6,26	4,43	6,38	0,00	6,22	6,45
PRODUTIVI DADE	2,50	2,74	4,81	8,28	6,22	0,00	7,09
TXIGPM	7,53	7,60	6,14	5,45	6,45	7,09	0,00

Fonte: Resultados da pesquisa.

Nota: Quantidade total produzido de grãos de soja brasileira (PRODSOJA), Quantidade total de importação de grãos de soja brasileira pela China (IMPCHINA), Taxa de Câmbio (TXCAMBIO), Taxa básica de juros SELIC (TXSELIC), Taxa de Desemprego (TXDESEMP), Produtividade calculada pela razão da quantidade total da produção de soja e área colhida em hectares (PRODUTIVIDADE), Índice Geral de Preços do Mercado (TXIGPM).

Baseando-se nas distâncias do Quadro 8 foram possíveis a elaboração dos gráficos do passo (Figura 2) e do dendograma (Figura 3) em que foi indicado os mesmos agrupamentos apresentados na Figura 1. Na Figura 2, o gráfico do passo, indica a linha de corte anexada ao dendograma onde ocorreu o primeiro maior salto entre as distâncias (8,2), que melhor indica os agrupamentos entre as variáveis.

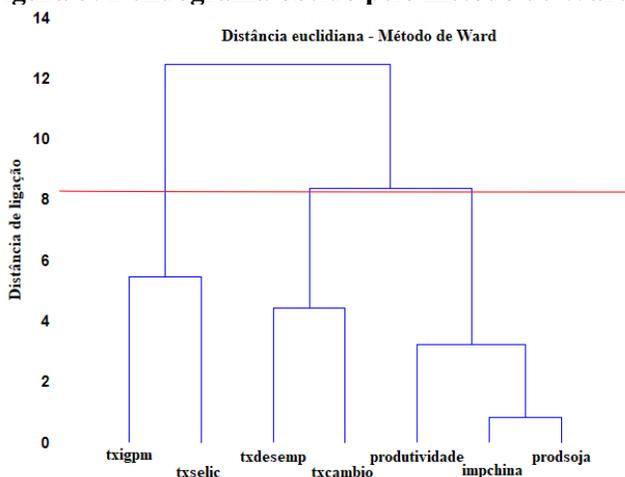
Figura 2: Gráfico das distâncias de ligação

Fonte: Resultados da pesquisa.

No dendrograma (Figura 3), evidencia a formação de três agrupamentos, baseados no corte feito na maior distância apresentada na Figura 2. Os coeficientes de distância são reescalados para valores entre 0 e 14. Notam-se agrupamentos similares com o Quadro 6 de ACP, que indicam as contribuições das variáveis com os valores calculados baseando-se na matriz de correlação.

O primeiro agrupamento do dendrograma, da direita para a esquerda, é composto pelas variáveis relacionadas à produtividade (PRODUTIVIDADE), importação de grãos de soja brasileira em milhares de toneladas pela China (IMPCHINA) e a produção brasileira em milhares de toneladas de grãos de soja (PRODSOJA). O segundo agrupamento, com um valor de escala pouco maior que quatro, é composto pelas variáveis referentes à taxa de desemprego (TXDESEMP) e a taxa câmbio (TXCAMBIO). E o último agrupamento é formado pelas variáveis de ordem econômica, taxa IGPM (TXIGPM) e a taxa SELIC (TXSELIC), próximos ao valor seis de escala.

É possível observar que o dendrograma (Figura 3) e o biplot poligonal (Figura 1) apresentaram o mesmo conjunto de grupos para as sete variáveis propostas. De mesmo modo que na ACP, as variáveis PRODSOJA, IMPCHINA e PRODUTIVIDADE foram as mais representativas no estudo. O acréscimo de mais fatores de produção na variável PRODUTIVIDADE, como por exemplo, mão-de-obra e capital, poderia acrisolar a análise, assim como o aprofundamento do comércio exterior na cultura da soja.

Figura 3: Dendograma obtido pelo método de Ward.

Fonte: Resultados da pesquisa.

Nota: Dendograma para as variáveis (Produção de Soja em milhares de toneladas de grãos -PRODSOJA, Taxa Selic -TXSELIC, Taxa de Câmbio - TXCAMBIO, Taxa IGPM - TXIGPM, Taxa de Desemprego - TXDESEMP, Importação da China em milhares de toneladas de grãos – IMPCHINA e Produtividade – PRODUTIVIDADE) padronizadas.

Conclusão

A destacada relevância das exportações de soja brasileira para a nossa economia e para a dinâmica do comércio internacional de *commodities* torna importante a avaliação dos fatores determinantes da produção deste grão no Brasil. Embora muito aplicados nas mais diversas áreas, os métodos multivariados de ACP e de AC são pouco explorados na economia, não tendo sido encontrado nenhum trabalho que avaliasse os fatores determinantes da produção de soja com base em tais metodologias.

Por meio de tal esforço, foi possível validar a importância dos métodos multivariados, bem como analisar de forma estatisticamente significativa a produção de soja no Brasil com respeito às variáveis de ordem econômica e monetária e a quantidade total de importação de grãos de soja brasileira pela China, referente aos anos de 1996 a 2017. Ambos os métodos evidenciaram a formação de três agrupamentos. A metodologia de ACP foi eficiente na redução do número de variáveis, com o mesmo grau de confiabilidade do conjunto de variáveis originais, resultando em duas componentes principais em que os autovalores somaram 75,42% da variação total. Também foi possível observar que a ACP obteve agrupamentos similares apresentados pelo dendrograma. Dessa forma, foi possível identificar que as variáveis quantidade total de importação de grãos de soja brasileira da China, produtividade e a taxa SELIC foram as que apresentaram relação mais significativa com a produção de soja, apesar de que

tratamentos localizados no lado esquerdo do biplot são questionáveis, como no caso da taxa SELIC. Além disso, foi possível verificar que a análise biplot permite uma análise eficiente para agrupamentos, auxiliando o pesquisador por meio da análise visual. Como proposta para novas pesquisas, o acréscimo de mais fatores associados à produtividade, tais quais mão-de-obra, capital e variáveis climáticas, assim como o aprofundamento do estudo a respeito do comércio exterior da cultura da soja.

Referências

AKINWALE, RO, FAKOREDE, MAB, BADU-APRAKU, B., & OLUWARANTI. Assessing the usefulness of GGE biplot as a statistical tool for plant breeders and agronomists. **Cereal research communications**, v. 42, n. 3, p. 534-546, 2014. DOI: <https://doi.org/10.1556/crc.42.2014.3.16>

Banco Central do Brasil. **Cotações e Boletins**. Disponível em: <https://www4.bcb.gov.br/pec/taxas/port/ptaxnpsq.asp?frame=1>. Acesso em: 5 mar. 2020.

BARROSO, Lúcia P.; ARTES, Rinaldo. **Análise multivariada**. Lavras: Ufla, 2003.

COMEX, STAT. **Ministério da Indústria, Comércio Exterior e Serviços**. Disponível em: <http://comexstat.mdic.gov.br/pt/comex-vis>. Acesso em: 26 out. 2020.

CONAB. **Perspectivas para a agropecuária**: volume 7, safra 2019/2020. Brasília, DF: CONAB, 2019. 102 p. Disponível em: https://www.conab.gov.br/perspectivas-para-a-agropecuaria/item/download/28825_2ed3fc3b5b25a350206d276620cf1c85. Acesso em: 30 jun. 2020.

COSTA, G. O. M. **A evolução do comércio entre Brasil e China: uma análise das relações comerciais bilaterais sino-brasileiras**. 2015. 32 f. Trabalho de conclusão de curso (Bacharelado - Ciências Econômicas) - Universidade Estadual Paulista Julio de Mesquita Filho, Faculdade de Ciências e Letras, Araraquara.

COSTA NETO, Pedro R. et al. Produção de biocombustível alternativo ao óleo diesel através da transesterificação de óleo de soja usado em frituras. **Química nova**, v. 23, n. 4, p. 531-537, 2000.

DOI: <http://dx.doi.org/10.1590/S0100-40422000000400017>

DE LIMA FILHO, Rafael Ribeiro; DA SILVA, Alex Santos Lopes; AGUIAR, Gustavo Adolpho Maranhão. Balanço e expectativas para a safra 2016/17. **AgroANALYSIS**, v. 36, n. 12, p. 21-23, 2016.

FERREIRA, D. F. **Estatística multivariada**. 3. ed. Lavras: UFLA, 2018. 624 p.

FIGUEIREDO, A. M.; SILVA, T. A. Exportação brasileira de soja em grãos: evolução e considerações sobre seus determinantes para o período de 1980–2001. **Organizações Rurais e Agroindustriais/Rural and Agro-Industrial Organizations**, v. 6, n. 1511-2016-131211, p. 81-91, 2004.

DOI: <http://dx.doi.org/10.22004/ag.econ.43675>

GABRIEL, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. **Biometrika**, v. 58, n. 3, p.453-467.

DOI: <https://doi.org/10.1093/biomet/58.3.453>

GOTELLI, N. J.; ELLISON, A. M. **Princípios de estatística em ecologia**. Porto Alegre: Artmed Editora, 2016.

HAIR, J. F., BLACK, W. C., BABIN, B. J., ANDERSON, R. E., & TATHAM, R. L. **Multivariate data analysis**: Pearson new international edition. Essex: Pearson Education Limited, 2014.

HERIG, Luiz. T. C. **O modelo das componentes principais: apresentação e interpretações**. Rio de Janeiro: UFRJ, 222p, 1979.

HIRAKURI, M. H.; LAZZAROTTO, J. J. O agronegócio da soja nos contextos mundial e brasileiro. **Embrapa Soja-Documentos (INFOTECA-E)**, Londrina: Embrapa Soja, 2014.

HIRAKURI, M. H.; CONTE, O.; PRANDO, A. M.; CASTRO, C. de; BALBINOT JUNIOR, A. A. Diagnóstico da produção de soja na macrorregião sojícola 4. **Embrapa Soja-Documentos (INFOTECA-E)**, 2019.

HOFFMANN, Rodolfo; PRONTO, L. **Análise estatística de relações lineares e não lineares**. São Paulo: LP-Books, 2011.

DOI: <https://doi.org/10.11606/9788592105716>

HONGYU, K.; SANDANIELO, V.L.M.; OLIVEIRA JUNIOR, G.J. Análise de Componentes Principais: Resumo Teórico, Aplicação e Interpretação. **E&S - Engineering and Science**, v.1, p.83-90, 2015.

DOI: <https://doi.org/10.18607/ES201653398>

HOTELLING, H. Analysis of a complex of statistical variables into Principal Components. *Jour. Educ. Psych.*, 24, 417-441, 498-520. **The Journal of Educational Psychology**, v. 24, p. 417–441, 1933.

DOI: <https://doi.org/10.1037/h0071325>

INSTITUTO BRASIEIRO DE GEOGRAFIA E ESTATÍSTIA (IBGE). **Painel de Indicadores**. Disponível em: <https://www.ibge.gov.br/indicadores>. Acesso em: 22 mar. 2020.

KAISER, Henry F. The application of electronic computers to factor analysis. **Educational and psychological measurement**, v. 20, n. 1, p. 141-151, 1960.

DOI: <https://doi.org/10.1177/001316446002000116>

KICH, Taciane Graciele Fanck; CORONEL, Daniel Arruda; VIEIRA, Kelmara Mendes. Determinantes da balança comercial do agronegócio brasileiro: análise da influência das variáveis macroeconômicas no período de 1997 a 2009. **Informações Econômicas**, v. 42, n. 4, p. 55-67, 2012.

LE MOS, Mario Luiz Freitas *et al.* **Agregação de valor na cadeia de soja**. Estudos Setoriais do BNDES, n. 46, p.167-217. 2017.

MANLY, Bryan. J. F., **Multivariate Statistical Methods: A primer**. Second edition. London: Chapman & Hall, 215p. 1994.

MENEZES, Gabrielito Rauter; FERNANDEZ, Rodrigo Nobre. Pass-through da taxa de câmbio e índices de preços: uma análise para a economia brasileira (1999-2011). **Perspectiva Econômica**, v. 9, n. 1, p. 31-42, 2013.

MINGOTI, S. A. **Análise de dados através de métodos estatística multivariada: uma abordagem aplicada**. Belo Horizonte: UFMG, 295p, 2007.

NAJBERG, S.; PEREIRA, R. de O. Novas estimativas do modelo de geração de empregos do BNDES. *Sinopse Econômica*, n. 133. Mar. 2004.

OLIVEIRA, Thamyres Alves Sena; DE OLIVEIRA NETO, Odilon José; SILVA, Edson Arlindo. Influência de variáveis socioeconômicas na produção brasileira de grãos. **Desafio Online**, v. 6, n. 3, 2018.

REIS, Elizabeth. **Estatística multivariada aplicada**. Lisboa: Edições Sílabo, 344p, 2001.

ROESSING, A. C.; LAZZAROTTO, J. J. **Criação de empregos pelo complexo agroindustrial da soja**. Londrina: Embrapa Soja, Documentos 233. 2004. 50p.

PEARSON, K. On lines and planes of closest fit to systems of points in space. **Philosophical Magazine**, v. 2, p. 559-572, 1901.

DOI: <https://doi.org/10.1080/14786440109462720>

QUADROS, Salomão. Os IGPs em 2002. **Revista Conjuntura Econômica**, v. 57, n. 2, p. 14-19, 2003.

RAMOS, C. M., Pizaia, M. G., Caldarelli, C. E., & da Camara, M. R. G. Competitividade e inserção da soja brasileira no mercado internacional. **Revista de Ciências Agrárias**, v. 43, n. spe1, p. 74-85, 2020.

DOI: <https://doi.org/10.19084/rca.19022>

ROMESBURG, H. C. **Cluster analysis for researchers**. North Carolina, EUA: Lulu.com, 334p. 2004.

SAITO, E. M. L.; OLIVEIRA, M. C. N. de. **Uso de variável *Dummy* para medir influência de variáveis socioeconômicas e a importação chinesa de grãos de soja**. XV Jornada Acadêmica da Embrapa Soja Resumos. Londrina: Embrapa Soja, Documentos 429. p. 26-35. 2019.

SILVA, Tamires Lamon Gomes. **Método Estatístico de Análise de Cluster aplicado aos dados de uma Associação de Proteção Veicular**. 2013.

STATSOFT. **STATISTICA** (data analysis software system). v. 7. Tulsa, Oklahoma, USA: Statsoft, 2007.

SOUZA, K. Alberto de; BITTENCOURT, G. M. Avaliação do crescimento das exportações brasileiras de soja em grão. **Revista de Política Agrícola**, v. 28, n. 4, p. 48, 2020.

SOUZA, M. Oliveira de. **Determinantes da disponibilidade interna, do processamento doméstico e da oferta de exportação brasileira de soja em grãos**. Tese (Doutorado em Economia Aplicada) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2018.

DOI: <https://doi.org/10.11606/T.11.2018.tde-29062018-160202>

SOUZA, M. Oliveira de; BACCHI, M. R. P; ALVES, L. R. A. ANÁLISE DE FATORES QUE INFLUENCIAM O PROCESSAMENTO DE SOJA NO BRASIL. **Revista de Economia e Agronegócio**, v. 17, n. 3, p. 485-506, 2019.

DOI: <https://doi.org/10.25070/rea.v17i3.7957>

United States Department of Agriculture. Foreign Agricultural Services. **Oilseeds: world market and trade (USDA-FAS)**, dez. 2019. 38 p. Disponível em: <https://downloads.usda.library.cornell.edu/usdaemis/files/tx31qh68h/n296xd515/w0892s45d/oilseeds.pdf>. Acesso em: 30 jun. 2020.

WARD, J. Hierarchical grouping to optimize an objective function. **Journal of American Statistical Association**, v.58, p.236-244, 1963.

DOI: <http://doi.org/10.1080/01621459.1963.10500845>

YAN, W.K.; RAJCAN, I. Biplot analysis of test sites and trait relations of soybean in Ontario. **Crop Science**, v.42, p.1120, 2002.

DOI: <https://doi.org/10.2135/cropsci2002.1100>

YAN, W. **Crop variety trials: data management and analysis**. West Sussex: J. Wiley, 2014. 351p.

DOI: <http://doi.org/10.1002/9781118688571>

YAN, W.; KANG, M.S. **GGE biplot analysis**: a graphical tool for breeders, geneticists, and agronomists. Boca Raton: CRC Press, 2003. 271p.

YAN, W; TINKER, Nicholas A. Biplot analysis of multi-environment trial data: Principles and applications. **Canadian journal of plant science**, v. 86, n. 3, p. 623-645, 2006.

DOI: <https://doi.org/10.4141/P05-169>