**ANÁLISE DA COMPONENTE PRINCIPAL DE INDICADORES DE DESEMPENHO EMPRESARIAL POR DECOMPOSIÇÃO DE VALORES SINGULARES**

 **Joaquim Eduardo de Moura Nicacio**

**Resumo:** O objetivo da pesquisa foi deduzir uma variável denominada componente principal sendo essa uma combinação linear de indicadores de desempenho empresarial do setor da economia Reflorestamento, Celulose e Papel. A transformação linear aplicada aos dados observados originais encontrou uma matriz ortogonal que maximizou a forma quadrática , que satisfaz o Teorema Espectral, com a restrição . O método aplicado foi a Decomposição de Valores Singulares através da utilização do ramo da matemática álgebra linear. Foi obtida uma única componente principal, contemplando 82% das informações sem nenhuma redundância entre as variáveis. A mudança de variáveis preservou a inércia dos valores observados e manteve as variáveis não correlacionadas.

**Palavras-chave:** Componente principal. Decomposição de Valores Singulares. Indicadores de desempenho empresarial.

***Abstract:*** *The research goal was to deduct a variable denominated main component, being this one a linear combination of entrepreneurial performance indicators from Reforestation, Celulosis and Paper economy. The linear transformation applied to the observed original data found an orthogonal matrix which maximized the quadratic form , that satisfies the Spectral Theorem, with the  restriction. Singular Value Decomposition was the method applied, by using the math branch of linear algebra. A single main component was obtained, contemplating 82% of the information without redundancies among the variables. The variables’ change preserved the inertia of observed values and kept the non related variables.*

***Key-words:*** *Main component. Singular value decomposition. Entrepreneurial performance indicators.*

**Classificação** **JEL: C02, M21**

**1. Introdução**

 Uma das medidas de desempenho econômico de empresas largamente empregada em estudos econômicos são os indicadores de balanço. Atendendo os aspectos legais impostos às empresas de capital aberto, tem-se as demonstrações contábeis auditadas por auditores independentes e, dentre eles, o Balanço Patrimonial.

 Um dos setores da economia que mais tem crescido independente das condições negativas de crescimento econômico que temos passado, é o Reflorestamento, Celulose e Papel que tem tido um desempenho digno de nota, principalmente, porque tem priorizado o desenvolvimento social e ambiental.

 Seu crescimento tem se dado na recuperação de áreas degradadas pela atividade da pecuária extensiva podendo reduzir bastante o custo de exploração madeireira. O emprego de tecnologia de ponta, a preocupação constante com a sustentabilidade, o acompanhamento do mercado globalizado necessário para uma política de exportação e a atenção às mudanças climáticas induzem o emprego de uma exploração de baixo carbono.

 Do ponto de vista da governança da empresa como um todo, pode-se utilizar os indicadores de desempenho empresarial presentes nas demonstrações contábeis e, de forma globalizada do setor. Esses mesmos indicadores são obtidos pela publicação dessas demonstrações em órgão institucional previsto em lei.

 Como cada empresa possui seus indicadores, é necessário entender a força de associação entre eles, ou seja: a covariância, bem como a dispersão em torno da média ou a variância. Como os dados observados podem ser colocados em uma matriz retangular, podemos colocá-la em forma de *desvio médio* ou *centrada* para obter uma matriz de variância-covariância.

 No entanto, à medida que se observa um maior número de empresas do mesmo setor e um maior número de indicadores, mais difícil se torna uma análise individualizada da relação empresa-indicadores.

 Diante dessa problemática, surge uma indagação: é possível reduzir a dimensionalidade dos dados multivariados, preservando ao mesmo tempo a maior quantidade de informação possível?

 A hipótese é que utilizando a álgebra linear e a programação linear, e aplicando uma transformação linear nos dados originais obtém-se um novo sistema de coordenadas, de tal forma que é obtido um novo conjunto de variáveis, as componentes principais, que são funções lineares das variáveis originais e, nessas, as mesmas não estão correlacionadas. Esse conjunto reduz o número de variáveis sem perda significativa de informação. O objetivo dessa transformação linear é encontrar uma matriz ortogonal  que maximiza  com a seguinte restrição . As colunas da matriz ortogonal P = U são vetores unitários associados aos valores próprios ou autovalores da matriz de covariância A. Esses vetores unitários da matriz de covariância são denominados componentes principais dos dados (na matriz de observações).

 A importância dessa análise reside na obtenção de um conjunto de novas variáveis, as componentes principais que são combinações lineares, que detêm o máximo de informação sem que essa seja redundante.

**2. Fundamentação teórica**

 As matrizes simétricas têm um papel relevante em aplicações na economia, contabilidade, engenharia, física, estatística, sensoriamento remoto, entre outras. A teoria correspondente é farta e depende, de maneira essencial, da técnica de diagonalização, tanto é verdade que se tem o seguinte teorema: *Uma matriz*  *é diagonalizável por matriz ortogonal se e somente se A é simétrica*, isto é: *D* = *.*

 Uma das aplicações de matrizes simétricas é na análise da componente principal, que é uma maneira efetiva de suprimir informação redundante, e pode se utilizar da matriz de variância covariância que é uma matriz quadrada e simétrica.

 A revista Globo Rural, publicada pela editora Globo, apresenta em seu 12º Anuário do Agronegócio as 500 maiores empresas do setor e, no caso do setor de *reflorestamento, papel e celulose,* apresenta as 10 empresas melhores do setor com respectivos indicadores de desempenho retirado das suas respectivas demonstrações contábeis. (Revista Globo Rural, 2016). No entanto, nesse artigo foram consideradas apenas nove empresas por motivos técnicos.

 Dos indicadores de desempenho foram destacados para uma análise em componentes principais cinco indicadores definidos pelos redatores do anuário que são:

* Rentabilidade do PL: 
* Margem Líquida: 
* Margem da Atividade: 
* Evolução do Ativo: 
* Evolução da Receita Líquida: .

 Considerando que os indicadores estão em uma mesma unidade de medida e devido a isso, pode haver comparação entre eles, o desenvolvimento da técnica das componentes principais se dará com o emprego da matriz de variância-covariância que é uma matriz quadrada e simétrica.

 O cálculo da inércia da nuvem de pontos representada pelos cinco indicadores de desempenho, das nove empresas permanece o mesmo quando forem calculados os *scores*, isto é: as novas coordenadas desses pontos-amostras.

 Entre os vários sistemas lineares da matemática, o da forma onde é um escalar (autovalor) e **x** (auto vetor), é um deles. Se prestarmos atenção, veremos que são de fato sistemas lineares homogêneos, pois, podem ser reescritos como , ou inserindo uma matriz identidade  e fatorando, como . Dado o sistema

 

Pode ser escrito em notação matricial como: . Pode ainda ser reescrito como: .

 Quando o sistema  tal que  onde *a*, *b*, *c* e *d* são conhecidos, o problema é determinar para quais valores de  esse sistema tem uma solução não trivial. Caso haja um determinado valor então, este é denominado *autovalor* de *A*. Se é um autovalor de *A*, então cada solução não trivial de  é chamada de *autovetor* de *A* associado ao autovalor . O sistema terá uma solução não trivial se 

1) Chama-se *vetor próprio* de uma transformação linear *f*, representada pela matriz T, a todo vetor não nulo X   tal que, para  real: ou

2) Se X ≠ 0 é vetor próprio da transformação linear *f*, representada pela matriz T, o número real  tal que TX = X é denominado *valor próprio* de *f.* (STEINBRUCH, 1975, p. 359).

Formas quadráticas são funções da forma , na variável  onde todos os expoentes são iguais a 2. Também se tem forma quadrática com *termo com produto misto* que contém variáveis diferentes como  onde a soma dos expoentes do termo do produto misto é igual a 2. Em notação matricial essa forma pode ser expressa da seguinte maneira:

  **(1)**

 Uma forma quadrática no plano pode ser representada por uma matriz simétrica que associa o vetor referido à base canônica  gera o polinômio homogêneo do 2º grau em *x* e *y* chamado *forma quadrática no plano.* Em notação matricial tem-se:

. Dessa forma a cada vetor  corresponde um número real

  **(2)**

 Podemos ainda eliminar o termo misto simplificando essa forma quadrática através de uma conveniente diagonalização. A diagonalização da forma quadrática acima pode ser realizada, pois sempre existe uma base ortonormal  em relação à qual a matriz de *Q* é diagonal.

 Considerando a matriz M onde *B* ≠ 0 podemos verificar que existem dois autovalores  e  reais e distintos. Achemos o determinante de M para obter a equação característica de  ou . Aplicando a fórmula de Bhaskara obtém-se:

 e  pois > 0.

 Como a base canônica *S*, origem do sistema de eixos, tem os vetores  e  que são perpendiculares entre si, isto é:  e tem comprimento unitário  podemos escrever: 

A toda matriz quadrada *A* sobre um corpo K está associado um escalar específico chamado *determinante de A*. Essa função determinante foi descoberta pela primeira vez na investigação de sistemas de equações lineares. Ela é uma ferramenta indispensável na investigação e obtenção das propriedades de um operador linear. (LIPSCHUTZ, 1973, p. 208).

 Considerando um exemplo numérico de uma matriz simétrica real  define no  a forma quadrática  que é uma hipérbole, sendo essa uma *cônica* conforme Figura 1.

Figura 1 - Hipérbole



 Fonte: Elaborado pelo autor.

Nos escritos de Pappus de Alexandria (350 – 290 a.C., aproximadamente), credita-se ao geômetra grego Aristeu (370 – 300 a.C.) a publicação do primeiro tratado sobre seções cônicas. Mais tarde, o astrônomo e matemático grego Apolônio de Perga recompilou e aprimorou os resultados conhecidos até então sobre o assunto na sua obra Seções Cônicas. Sabe-se pouco sobre a vida de Apolônio de Perga, sul da Ásia Menor. Sugere-se que viveu de 262 a 190 a.C. – parece ter-se considerado um cordial rival de Arquimedes. Supõe-se ter sido educado em Alexandria e por algum tempo ter ensinado em sua “Universidade”. Graças ao apoio de Lisímaco, general de Alexandre, transferiu-se para Pérgamo (donde a palavra pergaminho), onde havia uma biblioteca e uma “Universidade” só inferiores às de Alexandria. (PERES, 2014, 10).

 O Teorema Espectral pode ser formulado em termos de matrizes da seguinte forma: “Se a matriz  é simétrica, então existe uma matriz ortogonal  tal que é uma matriz diagonal que são os autovalores de ”. (LIMA, 2016, p. 211).

 A generalização das *cônicas* está na definição: uma *cônica* em  é um conjunto de pontos cujas coordenadas, em relação à base canônica, satisfazem a equação

  **(3)**

com *A* ou *B* ou *C* ≠ 0. Sua expressão em termos de matrizes é:

  **(4)**

 Uma das aplicações matemáticas importantes relacionadas com formas quadráticas é a maximização de com a seguinte restrição  onde essa forma quadrática tem um valor máximo de (o maior autovalor) e um valor mínimo de  (o menor autovalor).

 Uma das aplicações das formas quadráticas é na Geometria Analítica *n*-dimensional. Um subconjunto  chama-se uma *quádrica central* quando existe uma forma quadrática  tal que é definido pela equação .

 Se para , isto significa que é o conjunto dos pontos  tais que .

As mais simples das curvas seguintes são aquelas definidas por equações quádricas nas quais não mais que duas variáveis são sempre multiplicadas entre si. Assim, elevar uma variável ao quadrado, ou multiplicar duas variáveis diferentes, é permitido, mas elevar uma variável ao cubo, ou multiplicar uma variável pelo quadrado da outra, é estritamente proibido. (ELLENBERG, 2015, p. 365).

**3. Metodologia**

Utilizou-se cinco dos indicadores de balanço do setor Reflorestamento, celulose e papel de nove empresas dentre as dez consideradas as melhores pelo Anuário do Agronegócio 2016, publicado pela Globo Rural conforme Tabela 1.

 **Tabela 1 -** Empresas do setor reflorestamento, celulose e papel

 com seus indicadores de balanço

|  |  |
| --- | --- |
| Empresa | Indicadores de Balanço |
| RPL | ML | MA | EA | ERL |
| Fibria | 2.7 | 6.0 | 18.4 | 24.5 | 45.6 |
|  CMPC | -6.2 | -31.9 | 46.7 | 68.6 | 162.9 |
| Eldorado Brasil | 25.6 | 9.5 | 40.5 | 24.2 | 40.4 |
| Suzano Papel e Celulose | -10.1 | -9.2 | 30.9 | 0.5 | 42.6 |
| Klabin | -23.4 | -22.3 | 24.9 | 27.4 | 16.2 |
| Cenibra | 7.2 | 11.1 | 24.9 | 34.8 | 30.0 |
| Sguario Florestal | 10.3 | 40.5 | 38.7 | 6.5 | 31.9 |
| Fibria MS | 5.0 | 15.4 | 15.5 | 14.8 | 19.1 |
| Duratex Florestal | 6.5 | 25.4 | 32.7 | -10.8 | -10.7 |

 Fonte: Revista Globo Rural (2016).

 Os dados foram dispostos em uma matriz retangular  para permitir o início da análise das componentes principais com utilização de uma matriz de variância-covariância.

 O valor da *inércia* da nuvem de pontos é calculado pela distância euclidiana entre os cinco pontos-amostras (indicadores de desempenho) cujo somatório dessas distâncias dividido pelo número de empresas é igual à soma das dispersões dos cinco indicadores.

 Em seguida, foram determinados na matriz  em cada coluna, a sua média aritmética e deduzido de cada elemento e de cada coluna a média respectiva da coluna, obtendo uma *matriz retangular* denominada *X.*

 Para obter a matriz de *variância covariância* de uma amostra, foi efetuado o produto da matriz transposta pela matriz dos dados centrados e divido por *n -* 1, isto é:  obtendo dessa maneira uma matriz quadrada e simétrica.

 Obtida a matriz*A,* então pode-se determinar a equação característica  = **0** onde . Em seguida obteve-se suas raízes que são seus autovalores pela resolução do polinômio característico, após cálculo do determinante da equação característica.

 Uma vez obtidos os autovalores, estes são colocados em uma matriz diagonal onde o maior autovalor está na posição  e o último em . A contribuição de cada autovalor ou valor próprio para a variância total é obtida pela divisão de cada autovalor pela soma total deles, isto é: pelo traço da matriz diagonal.

 A obtenção dos autovalores também ocorre com a *Decomposição Em Valores Singulares* que é uma das fatorações mais úteis em álgebra linear aplicada.

A decomposição em valores singulares é baseada na seguinte propriedade da diagonalização usual que pode ser imitada para matrizes retangulares: os valores absolutos dos autovalores de uma matriz simétrica *A* medem as quantidades que *A* estica ou encurta certos valores (os autovetores). Se e , então . Se é o valor de maior módulo, então um autovetor unitário associado , identifica a direção na qual o efeito de esticar de é maior. Em outras palavras, o comprimento de é máximo quando , e . (LAY, 2007, p. 428).

 Pode-se mostrar que, para qualquer matriz simétrica *A*, o conjunto de todos os valores possíveis de  para  é um intervalo fechado na reta real. Os limites inferior e superior desse intervalo podem ser expressos por:  e .

**4. Resultados**

 A matriz de dados originais que irá permitir o cálculo da matriz de variância-covariância.



 A matriz das distâncias euclidianas entre as empresas é:



 A soma das distâncias é: 417.64 + 1110.02 + 288.67 + 253.19 + 248.49 + 131.60 +86.72 + 44.06 = 2.580,39. O valor da *inércia* é .

 A matriz *X* centrada é:



 A matriz de Variância/Covariância foi obtida pela multiplicação da matriz transposta centrada de *X* pela matriz centrada *X* multiplicada pela fração , isto é:sendo obtida a matriz quadrada e simétrica



Com base nessa matriz, foi determinada a equação característica . O determinante dessa equação forneceu o polinômio característico





 As raízes que são os autovalores são em ordem decrescente:



Seja Σ uma matriz não negativa (isto é, definida positiva ou semi-definida positiva) de ordem *p x p.* As raízes da equação determinantal são chamadas de valores próprios de . Essa equação tem *p* raízes reais não negativas Correspondentemente a cada raíz, existe um vetor coluna ***Pi,***, tal que que é chamado um vetor próprio. (RAO, 1964, p. 330).

 Os valores singulares são:



 O traço da matriz D é  que representa a variância total. O autovalor 3018.90 representa  e o autovalor 432.96 representa . Sob essas condições, apenas uma componente principal é capaz de explicar 82% da variância.

 Os auto vetores normalizados associados aos autovalores decrescentes correspondem as colunas da esquerda para a direita. Esses vetores foram obtidos na *Decomposição em Valores Singulares.*



 O vetor U é ortogonal, isto é:  onde .

; 



 Fazendo, 

Onde  o vetor associado ao maior autovalor  então,  é **3018.5**,diferindo por arredondamento na multiplicação. Essa é a otimização da forma quadrática  com a restrição .

 O valor da inércia considerando as novas coordenadas (*scores*) das amostras (empresas) é dado pela seguinte equação matricial  onde: F é a matriz das coordenadas das nove amostras sobre o eixo fatorial; X é a matriz dos dados (nove empresas e cinco indicadores) centrados; **U** é a matriz dos auto vetores (coeficientes).



 A soma das distâncias é: 417.64 + 1110.02 + 288.67 + 253.19 + 248.49 + 131.60 +86.72 + 44.06 = 2.580,39. O valor da *inércia* é . A inércia é preservada com as novas coordenadas. Logo a componente principal tem a seguinte equação:



 Como se pode observar a restrição é satisfeita pois:



**5. Conclusões**

 A decomposição em valores singulares é a principal ferramenta para efetuar a análise de componentes principais em aplicações práticas. Se X é uma matriz de observações n x m em forma de desvio médio (matriz centrada) e se  então A é a matriz de covariâncias. Os quadrados dos valores singulares de A são os *m* autovalores de *A* e os auto vetores singulares à direita de *A* são os componentes principais dos dados.

 O autovalor 3018.90 representa  e o autovalor 432.96 representa  logo, a diferença percentual é bastante expressiva, o que fez com que fosse considerada apenas uma única componente principal reduzindo drasticamente o número de variáveis, sem perda significativa de informação.

 Considerando uma análise de uma única empresa, os meses substituirão as empresas, e as variáveis podem permanecer como estas ou ampliadas. Dessa forma a análise poderá ocorrer anualmente ou como desejar a empresa. Isso dará uma contribuição muito valiosa para a gestão empresarial.

 A maximização de com a seguinte restrição  ocorreu uma vez que o maior autovalor 3018.90 representando 82% da informação foi o único a ser utilizado.

 O emprego da Álgebra Linear na dedução da componente principal permite um desenvolvimento de análise profundo, dando segurança na elaboração dos cálculos pois permite efetuar as validações de vários teoremas.

**Referências**

DE JESUS, G. D. S. *Aproximação Numérica de Valores Próprios de Matrizes Reais.*2012. 125 f. Dissertação (Mestre em Matemática e Aplicações) – Curso de Mestrado em Matemática, Universidade de Aveiro, Portugal. 2012.

ELLENBERG, J. *O poder do pensamento matemática:**a ciência de como não estar errado.*Tradução George Schlesinger; revisão técnica Samuel Jurkiewicz. Rio de Janeiro: Zahar, 2015.

LAY, D. C. *Álgebra Linear e suas aplicações.* Tradução Ricardo Camelier, Valéria de Magalhães Iório. 2. ed. Rio de Janeiro: LTC, 2007.

LIMA, E. L. *Álgebra Linear.*9. ed. Rio de Janeiro: IMPA, 2016.

LIPSCHUTZ, S. *Álgebra Linear.*Traduzido por Roberto Ribeiro Baldino. São Paulo: McGraw-Hill do Brasil, 1973.

LUZARDO, D. *et al.* Historia del Álgebra Lineal hasta los Albores del Siglo XX. *Divulgaciones Matemáticas*, v. 14, n. 2, p. 153-170, 2006.

ORDÓÑEZ, G. R. R*. Diseño e implementación de talleres para la enseñanza y aprendizaje del álgebra matricial y solución de sistemas de ecuaciones lineales con Scilab*. 2012. 205 f. Tese (Magister en Enseñanza de las Ciencias) – Magister en Enseñanza de Ciencias. Universidad Nacional de Colombia. Manizales. Colombia. 2012.

PERES, E. S. *Classificação de Cônicas e Quádricas em Função da Equação Algébrica.* 2014. 95 f. Trabalho de Conclusão de Curso (Mestre em Matemática) – Programa PROFMAT, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil. 2014.

RAO, C. R. The use and Interpretation of Principal Component Analysis in Applied Research. *Sankhya: The Indian Journal of Statistics.*Series A (1961-2002), v. 26, n. 4 (Dec., 1964), p. 329-358.

REDAÇÃO GLOBO RURAL. As 500 maiores empresas do setor. 12º Anuário do Agronegócio**.** *Revista Globo Rural.* São Paulo, n. 12, out. 2016.