

YURY VASCONCELLOS DA SILVA

ANÁLISE DE CORRESPONDÊNCIA: UMA ABORDAGEM GEOMÉTRICA

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2012

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

S586a
2012

Silva, Yury Vasconcellos da, 1976-

Análise de correspondência: uma abordagem geométrica /
Yury Vasconcellos da Silva. – Viçosa, MG, 2012.
xi, 140f. : il. (algumas col.) ; 29cm.

Orientador: Carlos Henrique Osório Silva.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 137-140

1. Análise de correspondência. 2. Tabelas de contingência.
3. Análise multivariada. 4. Cana de açúcar. I. Universidade
Federal de Viçosa. II. Título.

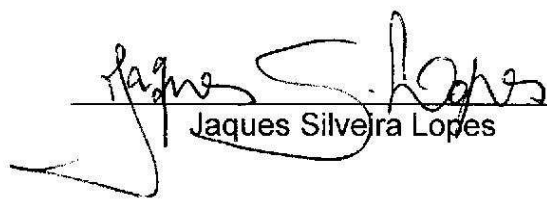
CDD 22. ed. 519.53

YURY VASCONCELLOS DA SILVA

ANÁLISE DE CORRESPONDÊNCIA: UMA ABORDAGEM GEOMÉTRICA

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

APROVADA: 28 de junho de 2012.


Jaques Silveira Lopes


Luiz Alexandre Peternelli


Adair José Regazzi
(Coorientador)


Gérson Rodrigues dos Santos
(Coorientador)


Carlos Henrique Osório Silva
(Orientador)

Dedico esta obra à minha família com
todo carinho e amor.

AGRADECIMENTOS

A Deus que me iluminou e me deu fé para acreditar que era possível.

À minha esposa Leila que foi o pilar de sustentação de nossa família durante minha dedicação para conseguir esta conquista.

Aos meus filhos Davi e Gabriela que são os grandes motivos da minha luta.

A meus pais, José Nazareno e Vilma que me deram ensinamentos que possibilitaram esta caminhada.

Aos meus tios Reginaldo e Gilda que sempre me incentivaram e apoiaram.

Aos meus primos que tenho como irmãos, Alice Carolina, Diego, Mayer e Mylliano.

Aos meus sogros Fabiano e Mercês que me acolheram com muito carinho durante esta caminhada.

À minha cunhada Maísa e ao seu esposo Marcelo pelo apoio que sempre me deram.

Ao amigo Pítias que sempre incentivou a realização deste sonho.

À amiga Gemma que lutou junto comigo para alcançar este objetivo, e que sem ela seria tudo seria muito mais difícil.

Aos mestres que me guiaram nesta empreitada, em especial ao meu orientador CHOS.

Aos amigos e companheiros do mestrado, em especial ao Gilson, Rafael e Rodrigo.

Ao meu sócio Ricardo que sempre me apoiou.

À Prof^a Jane, que me abriu as portas da Estatística.

“A mente que se abre a uma nova ideia
jamais voltará ao seu tamanho
original”. (Albert Einstein)

SUMÁRIO

| | |
|---|-----|
| LISTA DE FIGURAS | VII |
| LISTA DE TABELAS | IX |
| RESUMO | X |
| ABSTRACT | XI |
| 1 INTRODUÇÃO | 1 |
| 1.1 Histórico | 2 |
| 1.2 Estudos e aplicações da AC nos dias atuais..... | 5 |
| 1.3 Objetivos | 7 |
| 1.3.1 Objetivo geral | 7 |
| 1.3.2 Objetivos específicos | 7 |
| 2 REVISÃO BIBLIOGRÁFICA..... | 9 |
| 2.1 Fundamentos de álgebra linear..... | 9 |
| 2.1.1 Introdução à álgebra linear..... | 10 |
| 2.1.1.1 Vetores | 10 |
| 2.1.1.2 Matrizes | 13 |
| 2.1.1.3 Espaço vetorial | 19 |
| 2.1.1.4 Sistemas lineares..... | 22 |
| 2.1.2 Projeções ortogonais..... | 23 |
| 2.1.3 Autovalores e autovetores..... | 26 |
| 2.1.3.1 Transformações lineares..... | 27 |
| 2.1.3.2 Relação entre transformações lineares e autovalores e autovetores | 28 |
| 2.1.4 Diagonalização e decomposição de matrizes | 34 |
| 2.1.4.1 Diagonalização e decomposição de matrizes quadradas..... | 34 |
| 2.1.4.2 Diagonalização e decomposição de matrizes simétricas..... | 36 |
| 2.1.4.3 Decomposição em valores singulares..... | 44 |
| 2.1.5 Ajuste de quadrados mínimos..... | 50 |
| 2.1.5.1 Projeção ortogonal e erro quadrático | 51 |
| 2.1.6 Formas quadráticas..... | 54 |
| 2.1.7 Subespaço ótimo | 59 |
| 2.2 Análise de correspondência: metodologia e um exemplo introdutório . | 61 |
| 2.2.1 Tabela e matriz de dados..... | 62 |
| 2.2.2 Matriz de correspondência..... | 64 |
| 2.2.3 Massa de linha e coluna..... | 65 |
| 2.2.4 Perfil de linha e coluna..... | 65 |

| | | |
|----------|---|-----|
| 2.2.5 | A nuvem de pontos | 70 |
| 2.2.6 | Centro de gravidade ou centroide | 71 |
| 2.2.7 | Métrica euclidiana ponderada ou distância de qui-quadrado | 73 |
| 2.2.8 | Inércia | 79 |
| 2.2.9 | Relação entre inércia e a estatística do teste de qui-quadrado..... | 83 |
| 2.2.10 | Espaço transformado centrado no centroide..... | 84 |
| 2.2.11 | Redução de dimensionalidade | 94 |
| 2.2.12 | Relação entre linhas e colunas | 97 |
| 2.2.13 | As coordenadas nos eixos principais | 99 |
| 2.2.14 | Pontos suplementares..... | 109 |
| 2.2.15 | Interpretação da AC | 110 |
| 2.2.15.1 | Medidas de auxílio na interpretação..... | 110 |
| 2.2.15.2 | Mapas perceptuais..... | 113 |
| 3 | MATERIAL E MÉTODOS..... | 124 |
| 4 | RESULTADOS E DISCUSSÃO | 127 |
| | CONCLUSÃO | 136 |
| | REFERÊNCIAS BIBLIOGRÁFICAS..... | 137 |

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1 – Vetores equipolentes no \mathbb{R}^2 | 10 |
| Figura 2 – Regra do paralelogramo para a soma de vetores..... | 11 |
| Figura 3 – Visão geométrica das operações com vetores: (a) soma e subtração de vetores; (b) multiplicação por escalar | 12 |
| Figura 4 – Visão geométrica da norma de um vetor | 12 |
| Figura 5 – Vetores no \mathbb{R}^2 e respectivos versores. | 13 |
| Figura 6 – Exemplos de subespaços (a) do \mathbb{R}^2 : s e r ; (b) do \mathbb{R}^3 : t e α | 20 |
| Figura 7 – Visão geométrica de um exemplo de combinação linear..... | 21 |
| Figura 8 – Gráfico comparativo ilustrando a mudança de base do vetor w | 24 |
| Figura 9 – Projeções ortogonais de w sobre os vetores da base v_1 e v_2 | 24 |
| Figura 10 – Visão geométrica do produto interno entre os vetores u e v , tendo v norma 1..... | 26 |
| Figura 11 – Comparação da transformação linear $T(v) = Av = w$ com geração de autovetor (a) e sem a geração do mesmo (b)..... | 28 |
| Figura 12 – Projeção do vetor w sobre q_1, q_2 e q_3 | 44 |
| Figura 13 – Visualização de algumas possibilidades do vetor v com a restrição $v = 1$ | 46 |
| Figura 14 – Vetores que formam uma base em \mathbb{R}^2 e no \mathbb{R}^3 | 49 |
| Figura 15 – Visão geométrica de um ajuste de quadrados mínimos | 51 |
| Figura 16 – Projeções do vetor u sobre o plano α , sendo u a projeção ortogonal..... | 51 |
| Figura 17 – Projeção do vetor u no subespaço W | 52 |
| Figura 18 – Gráficos das formas quadráticas $f(x, y)$ | 56 |
| Figura 19 – Representação tri-dimensional dos perfis de linha sob dois ângulos de visão. | 71 |
| Figura 20 – Representação do perfil médio de linha (g) com os perfis de linha no espaço tridimensional | 73 |
| Figura 21 – Representação, sob 3 ângulos de visão, dos perfis de linha e o centroide no espaço original sob a métrica de qui-quadrado e no espaço transformado sob a métrica euclidiana | 86 |
| Figura 22: Translação da base para o centroide do espaço transformado..... | 87 |

| | |
|--|-----|
| Figura 23: Representação comparativa dos perfis h_i e x_i no espaço transformado..... | 90 |
| Figura 24: Representação eixos principais v_1, v_2, v_3 no espaço transformado | 96 |
| Figura 25 – Representação dos vetores m_1 e m_2 no espaço original | 107 |
| Figura 26 – Projeção dos resíduos dos perfis de linha $R - \mathbf{1}_l m_c^t$ sobre m_1 e m_2 | 108 |
| Figura 27 – Decomposição da distância do perfil de linha r_i pertencente ao \mathbb{R}^3 sobre os eixos a, b, c..... | 112 |
| Figura 28 – Mapa perceptual simétrico do exemplo ilustrativo da Tabela 2 ... | 120 |
| Figura 29 – Mapa perceptual assimétrico principal por linha do exemplo ilustrativo da Tabela 2..... | 122 |
| Figura 30 – Mapa perceptual assimétrico principal por coluna do exemplo ilustrativo da Tabela 2..... | 123 |
| Figura 31 – Mapa perceptual simétrico das características da cana-de-açúcar e peso da cana | 127 |
| Figura 32 – Mapa perceptual simétrico das características da cana-de-açúcar e peso da touceira destacando-se os pontos que devem ser interpretados pelo eixo 1 | 131 |
| Figura 33 – Mapa perceptual assimétrico principal por linha das características da cana-de-açúcar classificando-as de acordo com a produção | 132 |
| Figura 34 – Mapa perceptual assimétrico principal por linha agrupando os perfis de linha de acordo com o número de colmos e, conseqüentemente, pelo peso de touceira | 134 |
| Figura 35 – Mapa perceptual principal por linha de características da cana e peso da touceira com a projeção dos pontos suplementares das famílias da cana | 135 |

LISTA DE TABELAS

| | |
|--|-----|
| Tabela 1 – Estrutura geral de uma tabela de contingência utilizada na AC simples..... | 62 |
| Tabela 2 – Exemplo ilustrativo: tabela de contingência com as frequências observadas das categorias de duas variáveis A e B..... | 63 |
| Tabela 3 – Frequências absolutas do uso de drogas ilícitas e escolaridade | 68 |
| Tabela 4 – Frequências relativas do uso de drogas ilícitas e escolaridade | 68 |
| Tabela 5 – Frequências relativas do uso de drogas ilícitas e escolaridade | 69 |
| Tabela 6 – Critério de categorização das variáveis explicativas do experimento com cana-de-açúcar | 125 |
| Tabela 7 – Decomposição da inércia sobre os eixos principais..... | 128 |
| Tabela 8 – Medidas descritivas da AC sobre o perfil de coluna multiplicadas por mil. | 128 |
| Tabela 9 - Medidas descritivas da AC sobre o perfil de linha multiplicadas por mil..... | 130 |
| Tabela 10 – Classificação dos perfis de linha quanto às faixas de peso da touceira com ordenação crescente..... | 133 |

RESUMO

SILVA, Yury Vasconcellos da, M.Sc., Universidade Federal de Viçosa, junho de 2012. **Análise de correspondência: uma abordagem geométrica.** Orientador: Carlos Henrique Osório Silva. Coorientadores: Adair José Regazzi e Gérson Rodrigues dos Santos.

A análise de correspondência é uma técnica estatística multivariada que permite realizar a análise simultânea de diferentes variáveis categóricas. Possui aspecto simples e intuitivo na apresentação de resultados por meio dos mapas perceptuais. Apesar da análise de correspondência estar disponível em diversos *softwares*, incluindo o sistema livre R, o principal problema que contribui para o desconhecimento do potencial desta técnica e de sua consequente difusão é a existência de diversos textos especializados de difícil compreensão, por apresentarem um alto nível de algebrismo, abstração e por omitirem passos elucidativos importantes. Por outro lado, a leitura pura e simples das técnicas de interpretação se torna desprovida de sentido convincente, que pode levar a equívocos e insegurança nas interpretações e impede até mesmo de tirar o maior proveito possível dos resultados. Assim, o objetivo deste trabalho foi elaborar um texto elucidativo que mostrasse os detalhes matemáticos de todas as etapas da análise de correspondência, de forma aplicada e com ênfase na construção geométrica. Uma aplicação desta técnica com dados de um experimento na área de melhoramento genético da cana-de-açúcar é apresentada na qual constatou-se que o número de colmos é um fator determinante para antecipar a categoria de peso da touceira. Esta informação é de fundamental importância, uma vez que pode facilitar os aspectos operacionais na seleção de famílias para o melhoramento genético, visto que o pesquisador não necessitará esperar a época da colheita para selecioná-las, pois poderá fazê-la previamente se baseando naquelas com os maiores números de colmos. Portanto concluiu-se que esta técnica pode ser promissora em ciências agrárias. Apesar da leitura do presente trabalho exigir tempo e disposição, o mesmo contribui para um melhor entendimento desta técnica, principalmente para um leitor não muito afeito aos desenvolvimentos algébricos, pois ele fornece uma visão completa do passo a passo da análise de correspondência simples através da visão geométrica e do exemplo ilustrativo.

ABSTRACT

SILVA, Yury Vasconcellos da, M.Sc., Universidade Federal de Viçosa, June, 2012. **Correspondence analysis: a geometric approach.** Advisor: Carlos Henrique Osório Silva. Co-Advisors: Adair José Regazzi and Gérson Rodrigues dos Santos.

Correspondence analysis is a multivariate statistics technique that allows the simultaneous analysis of different categorical variables. It has a simple and intuitive aspect in the presentation of results through perceptual maps. Although correspondence analysis is available in several softwares, including free system R, the main problem that contributes to the lack of knowledge about the potential of this technique and its subsequent diffusion is the existence of several literatures of difficult comprehension, as they present linear algebra with a high level of abstraction and, additionally, for omitting important clarifying steps; on the other hand, the pure and simple reading of interpretation techniques simply become steps lacking in convincing meaning, which may lead to misunderstandings and insecurity in the interpretations and even takes away from the highest possible results. Thus, the aim of this study was to elaborate a clarifying text that showed the mathematical details of all the steps of the correspondence analysis in the form-applied method and with emphasis on its geometric construction. One of the applications of this technique with data from an experiment in the area of genetic improvement of the sugar cane and it allowed important and practical conclusions, such as, the number of stalks is a key factor to anticipate the weight of the clump. This information is of the highest importance, since it can facilitate the operational aspects of selection of families for genetic enhancing, and the researcher won't need to wait for harvest season to select the families with the best clump weight traits, as they will be able to select the clump based on largest number of stalks. Therefore, it was concluded that this technique can be promising in the agrarian sciences. Even though the reading of this study takes time and disposal, it contributes to a better understanding of this technique, especially for a layman in algebraic developments, since it provides a complete vision of the step by step of the correspondence analysis through geometric vision and illustrative example.

1 INTRODUÇÃO

A análise de correspondência (AC) é uma técnica estatística multivariada que permite realizar a análise simultânea de diferentes variáveis categóricas. Estas variáveis categorizadas podem possuir níveis de mensuração nominal e ordinal, ou ainda podem ser provenientes de variáveis quantitativas discretas ou contínuas. A AC é essencialmente uma análise exploratória e descritiva dos dados que permite uma análise gráfica das inter-relações ou correspondências entre as variáveis. Possui aplicabilidade nas mais diversas áreas de conhecimento, tais como: ciências sociais (economia, administração, sociologia, demografia), humanas (psicologia, história), ciências biológicas e da saúde, como também nas áreas agrárias. Especialmente na área de ciências sociais e humanas esta técnica demonstra ser de grande valia, visto que é comum a utilização de questionários com variáveis categóricas para a coleta de informações.

Segundo Favero et al. (2009), denomina-se análise de correspondência simples quando a AC é aplicada para duas variáveis. Quando AC é utilizada para três ou mais variáveis, a técnica é uma generalização da AC simples e recebe a denominação de análise de correspondência múltipla (ACM), também denominada de *Optimal Scaling* ou pela sigla HOMALS, cujo significado em inglês é *homogeneity analysis via alternating least squares*, (HILDEBRAND; MÜLLER-FUNK, 2012).

A análise de correspondência simples possui aplicação relacionada às tabelas de frequência de dupla entrada. AC permite elucidar o grau de relacionamento entre as categorias das linhas, ou entre as categorias das colunas, ou ainda, entre o conjunto de categorias das linhas com o conjunto de categorias das colunas (CZERMAINSKI, 2004).

É comum a utilização do teste de qui-quadrado para verificar a existência ou não de relação de dependência entre duas variáveis categóricas, porém esta técnica não demonstra em quais categorias existe essa dependência, como também, não demonstra sua intensidade.

Em contraposição, na AC a visualização destes padrões de relacionamento se dá através da interpretação de gráficos bidimensionais, denominados gráficos perceptuais, onde são plotados os pontos resultantes da redução da dimensionalidade do conjunto de dados. Neles, é possível observar

se as variáveis de interesse distanciam-se do pressuposto de independência convergindo para possíveis associações existentes entre elas, sendo possível ainda verificar como se dá esta associação, uma vez que níveis das variáveis de linha e de coluna assumem posições nos gráficos de acordo com suas associações ou com suas similaridades (GREENACRE, 1984; BENZÉCRI, 1992; GREENACRE, 2007). Ou seja, o mapa perceptual gerado na análise de correspondência demonstra, de maneira simples, a existência de dependência entre as variáveis, bem como quais categorias estão relacionadas.

Apesar de estar implementada em diversos softwares, um fator ainda limitante para o entendimento da AC por usuários (pesquisadores, estudantes e/ou profissionais) são seus cálculos matemáticos complexos provenientes da análise multivariada. Neste sentido, textos elucidativos e de fácil leitura, aliado ao uso de ferramentas computacionais torna-se indispensável para implementar esta técnica que, de acordo com Souza (2004), tem se tornado cada vez mais popular e reconhecida como poderosa ferramenta estatística de análise exploratória de dados.

1.1 Histórico

Para melhor compreender o desenvolvimento da análise de correspondência, pode-se dividir sua evolução histórica em três fases distintas: (i) desenvolvimento algébrico, (ii) desenvolvimento geométrico e sua (iii) expansão científica.

A análise de correspondência foi originalmente criada para análise de grandes tabelas de contingência envolvendo duas variáveis categóricas. Atualmente o seu uso se estende para tabelas de frequência com mais do que duas variáveis. Nyfjäll (2002a, p. 5), esclarece as diferenças entre tabelas de contingência e tabelas de frequências. Essencialmente, em uma tabela de frequência cada célula informa o número de vezes que determinada combinação de categorias ocorreu, sendo que, um objeto ou indivíduo pode pertencer ou escolher mais de uma categoria. Já em uma tabela de contingência cada célula informa o número de objetos ou indivíduos que pertencem ou escolheram aquela combinação de categorias.

Existem várias técnicas de análise aplicadas a tabelas de contingência, tais como modelos log-lineares, modelos logísticos, escalonamento ótimo,

análise de homogeneidade e a estatística qui-quadrado de Karl Pearson (1900), sendo esta última uma das mais importantes e conhecidas e que inspirou o desenvolvimento algébrico da análise de correspondência (PAMPLONA, 1998).

Em 1935 H. O. Hartley publicou sob seu nome em alemão – Hirschfeld – um artigo matemático descrevendo a formulação algébrica da “correlação” entre linhas e colunas em uma tabela de contingência. Pode-se atribuir a origem matemática da análise de correspondência a este artigo, apesar dos trabalhos de Ricardson and Kunder (1933) e Horst (1935) terem sugerido independentemente a mesma ideia na literatura psicométrica, sem o respectivo desenvolvimento matemático, sendo que Horst inclusive sugeriu o termo método das médias recíprocas (GREENACRE, 1984, p. 8).

Na década de 40, houve avanços significativos na formulação matemática da análise de correspondência. Segundo Greenacre (1984, p. 8, tradução nossa), Fisher (1940) derivou a mesma teoria de Hartley na forma de análise discriminante de uma tabela de contingência, aplicada a dados biométricos relativos à coloração dos olhos e dos cabelos de estudantes. Guttman (1941), independentemente, derivou o método construindo escalas para dados categóricos, sendo na essência a mesma teoria de Fisher, porém aplicada a dados psicométricos. Foi também Guttman que introduziu a forma geral para lidar com tabelas de contingência com mais do que duas variáveis, conhecido como dual (ou *optimal*) *scaling*, origem da análise de correspondência múltipla (SOUZA, 2004, p. 68).

Greenacre (1984) e Khattree e Naik (2000) citam também como sendo as principais contribuições históricas para o desenvolvimento da análise de correspondência, os trabalhos de Hotelling (1933), Thurstone (1935), Eckart and Young's (1936), Mosier (1946), Rao (1948), Burt (1950), Hayashi (1950, 1952, 1954, 1968), Williams (1952) e Bock (1960). Sendo assim, houveram inúmeras contribuições na técnica, o que acarretou no surgimento de diversos termos para esta técnica, tais como HOMALS, *Optimal Scaling* e *Dual Scaling*.

A segunda fase da evolução da análise de correspondência originou-se com um grupo de pesquisadores franceses liderados por Jean-Paul Benzécri, que formulou a forma geométrica da análise de correspondência como a conhecemos atualmente. Este grupo iniciou seus estudos durante a década de 60 na área de linguística, tendo seu desfecho em uma surpreendente obra de

1973, composta de dois volumes a saber: “*L’Analyse des Donnés*” (volume I) e “*L’Analyse des Correspondance*” (volume II).

Este grupo de pesquisadores franceses merece destaque por terem desenvolvido ideias e princípios que se estenderam bem mais além da área inicial de trabalho, a linguística, e por preconizarem muitas ideias inovadoras relacionadas à análise de dados, tais como um princípio defendido por Benzécri: “O modelo deve se ajustar aos dados, e não o contrário” (GREENACRE, 1984, p.10, tradução nossa). Este princípio é uma rejeição inicial à modelagem usual com modelos matemáticos probabilísticos, com suas pressuposições que quase sempre não são atendidas em sua plenitude.

Neste sentido, Benzécri desenvolveu técnicas geométricas de visualização dos dados na forma de pontos em um espaço multidimensional mostrados em gráficos bidimensionais. Outra particularidade de Benzécri e sua equipe era a utilização de uma notação algébrica extremamente rigorosa e poderosa. Ao mesmo tempo, esta linguagem matemática diferenciada, dificultava a comunicação e o entendimento da técnica de análise de correspondência pelos pesquisadores da escola anglo-americana. Khattree e Naik (2000, p. 444, tradução nossa) afirmam que a “análise de correspondência se tornou uma ferramenta popular entre os profissionais do resto do mundo após o lançamento do livro de Greenacre (1984)”.

Greenacre é sul africano e conheceu a metodologia de análise de correspondência durante seu doutorado na França, orientado pelo próprio Benzécri. Em 1980 após o término do doutorado, foi convidado a palestrar em uma conferencia internacional de métodos gráficos multidimensionais que ocorreu na Inglaterra e nesta oportunidade percebeu o grande interesse da comunidade científica pelo assunto, e então resolveu escrever um livro lançado em 1984 sob o título: “*Theory and Applications of Correspondence Analysis*”.

A terceira fase da evolução da análise de correspondência inicia-se após a publicação de Greenacre (1984), que além de traduzir a metodologia para o inglês adaptou a notação algébrica diferenciada de Benzécri para que pudesse ser entendida pelos pesquisadores e estudiosos ao redor do mundo. Surgiram então numerosas publicações nesta área e conseqüentemente o aperfeiçoamento do algoritmo original além de adaptações e variantes, tais como (NAITO, 2007): análise canônica de correspondência simples ou parcial (Ter Braak, 1985, 1986 e 1988), análise de correspondências conjuntas

(Greenacre, 1988), análise de correspondência não simétrica (D'Ambra e Lauro, 1989), análise de correspondência com restrições (Takane, et al., 1991), análise de correspondência utilizando a distância de hellinger (Rao, 1995), análise de correspondência ordinal (Beh, 1997), análise de correspondência inversa (Groenen Van de Velden, 2002) e rotação em análise de correspondência (Van de Velden e Kiers, 2003).

Apesar de existirem inúmeras publicações a respeito da AC na língua inglesa, na literatura brasileira existem inúmeros trabalhos com a aplicação da AC, porém são escassos os trabalhos que expliquem os detalhes matemáticos e estatísticos relacionados a esta técnica, principalmente para pesquisadores de outras áreas que não a Estatística.

Neste contexto, o estímulo para realização do presente trabalho está pautado na necessidade de um texto elucidativo que mostre os detalhes matemáticos de forma aplicada.

1.2 Estudos e aplicações da AC nos dias atuais

Duas vertentes são comuns nos artigos envolvendo a AC: os estudos teóricos e as aplicações, sendo estas últimas mais comuns nas ciências sociais.

Os estudos teóricos envolvem tanto derivações da AC simples ou múltipla que atendam a certas configurações especiais de dados como também associações da AC a outras análises de agrupamentos com o objetivo de otimizar a interpretação de seus resultados.

Foi exatamente com esta intenção que Alves, Belderrain e Scarpel (2007) utilizaram a análise de agrupamentos conjuntamente com a AC num estudo referente à associação de variáveis relativas à saúde física e mental de idosos de uma comunidade. Neste estudo, os autores ressaltam, também, a importância de pesquisas que expõem a combinação de métodos como meio de aperfeiçoar resultados. Tal opinião é compartilhada por Silva Neto e Raposo (2010).

Na mesma linha de associação de métodos, pode-se ainda citar Guedes et al. (1999), que propõem em seu trabalho um procedimento que aplica a análise procrustes combinada com a AC para encontrar um ranking de importância para as colunas (atributos) de uma tabela de contingência. E ainda,

Silva et al. (2008), que fazem uso da análise de cluster para auxiliar nas interpretações dos mapas perceptuais obtidos pela AC em dados envolvendo as relações de percepção da alta administração das empresas produtoras de móveis da região Sul do Brasil quanto a ameaças encontradas em relação à Área de Livre Comércio das Américas (ALCA).

Avançando nos estudos teóricos, Greenacre e Fargas (2010) propõem uma variante da análise de correspondência, a análise de correspondência de dados brutos (não relativizados) em amostragem de aplicações ecológicas, nas quais é comum que sejam realizadas em áreas iguais ou volumes iguais. Souza, Bastos e Vieira (2010) utilizaram os resultados da AC múltipla para propor uma metodologia de escores ponderados em pesquisa de satisfação.

Beh (2004) trabalha com questões teóricas, práticas e computacionais da análise de correspondência simples e discute sua relação com os avanços da época – e ainda são válidos para hoje – nas exibições gráficas. Mais tarde apresenta outros resultados de estudos da AC, envolvendo viesamento e curtose (BEH, 2009).

Também na linha computacional, de Leeuw e Mair (2007) apresentam e exploram os recursos adicionais que o pacote ANACOR possui que não são oferecidos por outros pacotes de AC do R, apoiados em referencial teórico; Figueira (2004) ensina a usar o procedimento ANACOR do SPSS e a interpretar seus resultados.

Enfocando tópicos controversos da AC, pode-se citar Carvalho, Vieira e Moran (2002), que apresentam a AC como um método estatístico multivariado que pode ajudar na interpretação de mapas de produtividade, mas lembrando de que é inadequado seu uso indiscriminado como ferramenta de informação preditiva. Guedes e Ivanqui (1999) também afirmaram que as conclusões obtidas na AC não podem ser generalizadas para a população, embora seja o que tem sido praticado. Souza, Bastos e Vieira (2011) em seu estudo mostram que os resultados da AC podem ser seriamente afetados pela amostragem. E Greenacre (2011) rebate as críticas sobre a sensibilidade da AC em relação a objetos raros, mostrando que seus deslocamentos de baixo peso reduzem seus efeitos sobre as contribuições de tais objetos tanto para a determinação dos eixos principais quanto para a distância de qui-quadrado.

Apesar das críticas e algumas desconfianças, a AC vem ganhando reputação internacional como uma ferramenta poderosa para a análise de

tabelas de contingência (BEH, 2004). Outro fator que impele seu crescimento, é o uso cada vez maior de dados categóricos tanto no campo teórico como no prático

Na linha da aplicação prática, estudos em diversas áreas comprovam os bons resultados da AC. Jelihovschi e Ferraz (2010) confirmam a eficiência da análise de correspondência para explorar agrupamentos de fatores socioculturais, utilizando um questionário preenchido pelos vestibulandos da UESC de 2008. Aranha et al. (2004) usam a AC em dados de pesquisa de avaliação de saúde pública. Na área agrária, pode-se citar Lopes et al. (2010) que utilizam a AC para relacionar situação de fontes de águas subterrâneas e fatores que interferem na qualidade da água. Ghini e Zaroni (2001) aplicaram AC para definir agrupamentos de características que levam à supressividade do solo. Carvalho, Vieira e Moran (2002) utilizaram a AC para a definição de mapas de produtividade dos solos.

1.3 Objetivos

1.3.1 Objetivo geral

O objetivo geral deste trabalho é a apresentação de um estudo a respeito da AC, com uma aplicação na área agrária, expondo os detalhes do desenvolvimento algébrico, dos princípios matemáticos e estatísticos da fundamentação teórica desta metodologia, com enfoque na evolução geométrica de suas etapas, propiciando assim um texto que permita o entendimento da AC por profissionais de áreas aplicadas.

1.3.2 Objetivos específicos

- Apresentar um resumo de álgebra linear, que aborde os principais conceitos necessários ao entendimento da AC, utilizando pequenos exemplos ilustrativos;
- Apresentar o desenvolvimento algébrico de cada etapa da AC, detalhando seus cálculos e explicando-os por meio da interpretação geométrica, a partir de um exemplo ilustrativo;
- Aplicar AC a um conjunto de dados oriundo do Centro de Pesquisa e Melhoramento de Cana-de-Açúcar da Universidade Federal de Viçosa para

exemplificar o uso desta técnica como uma alternativa metodológica de análise estatística na área de ciências agrárias.

2 REVISÃO BIBLIOGRÁFICA

Como foi dito na introdução, a análise de correspondência requer conhecimentos matemáticos, que em sua maioria, não estão ao alcance de muitos usuários da estatística, principalmente aqueles cuja formação acadêmica não é da área de ciências exatas. Embora a aplicação desta técnica seja tremendamente facilitada pela disponibilidade de pacotes computacionais especializados, o entendimento dos métodos matemáticos utilizados é importante para uma adequada compreensão da técnica, de sua aplicabilidade, da interpretação dos resultados e de suas consequências, favorecendo assim uma análise mais segura e elucidativa dos dados.

Nesse sentido, nesta primeira subseção (item 2.1) serão apresentados os conceitos e definições de álgebra linear utilizados na análise de correspondência, tais como operações com vetores e matrizes, projeções ortogonais, transformações lineares, autovalores e autovetores, formas quadráticas, ajuste de mínimos quadrados, diagonalização e decomposição de matrizes, sendo este último de fundamental importância para análise de correspondência.

Na segunda subseção (item 2.2) serão apresentados todos os procedimentos algébricos necessários para realização de uma análise de correspondência, com os conceitos, fundamentação teórica e interpretação geométrica envolvidos em cada um de seus passos.

2.1 Fundamentos de álgebra linear

Toda a fundamentação teórica de álgebra linear apresentada neste trabalho está baseada nas obras dos seguintes autores: Boldrini et al. (1978), Lay (1999), Leon (1999), Poole (2004) e Matos e Amaral (2008). Em alguns casos, foram utilizados os exemplos destes autores os quais foram devidamente citados.

2.1.1 Introdução à álgebra linear

Nessa subseção será realizada uma rápida revisão dos aspectos elementares de álgebra linear. Portanto, caso o leitor não os conheça é recomendado o seu respectivo estudo nas obras já citadas anteriormente ou em qualquer literatura de álgebra linear básica.

Ao mesmo tempo, se o leitor já está familiarizado com operações básicas de vetores e matrizes, poderá ignorar esta subseção e adiantar sua leitura para o item 2.1.2 - projeções ortogonais ou outro tópico que não esteja tão familiarizado, tais como transformações lineares, autovalores e autovetores, diagonalização e decomposição de matrizes, ajustes de mínimos quadrados e formas quadráticas.

Aqueles que possuem domínio dos assuntos acima citados poderão iniciar sua leitura a partir do item 2.2 que aborda a AC propriamente dita.

2.1.1.1 Vetores

Um vetor v no \mathbb{R}^n é um arranjo de n elementos reais tais que $v = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$.

Geometricamente, v é um ente do \mathbb{R}^n com módulo, direção e sentido. Dois vetores são ditos equipolentes se têm o mesmo módulo, a mesma direção e o mesmo sentido, conforme ilustrado na Figura 1. Todos os vetores equipolentes podem ser representados por um único vetor, sendo este o que parte da origem do sistema cartesiano.

Define-se como v transposto o vetor linha $v^t = [x_1 \ x_2 \ \dots \ x_n]$.

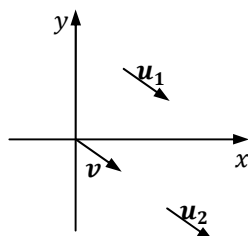


Figura 1 – Vetores equipolentes no \mathbb{R}^2

Operações com vetores

Considerando $\mathbf{u} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$, $\mathbf{v} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n$ e $k \in \mathbb{R}$, então:

(i) a soma de dois vetores \mathbf{u} e \mathbf{v} é dada por $\mathbf{u} + \mathbf{v} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$, cuja visão geométrica, no \mathbb{R}^2 , é apresentada na Figura 2;

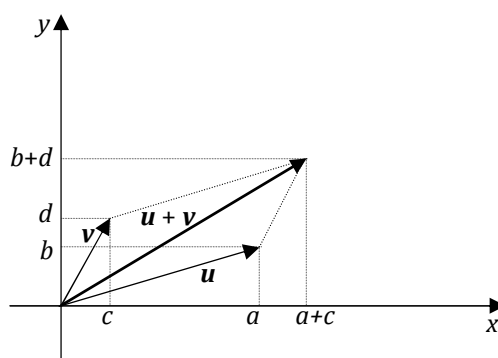


Figura 2 – Regra do paralelogramo para a soma de vetores

(ii) a multiplicação de um vetor por uma constante real (escalar), chamada de

multiplicação por escalar, é dada por $k\mathbf{u} = \begin{bmatrix} kx_1 \\ kx_2 \\ \vdots \\ kx_n \end{bmatrix}$;

(iii) a diferença entre dois vetores \mathbf{u} e \mathbf{v} é dada por

$$\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v}) = \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{bmatrix}.$$

A Figura 3 mostra a visão geométrica no \mathbb{R}^2 das operações com vetores acima descritas.

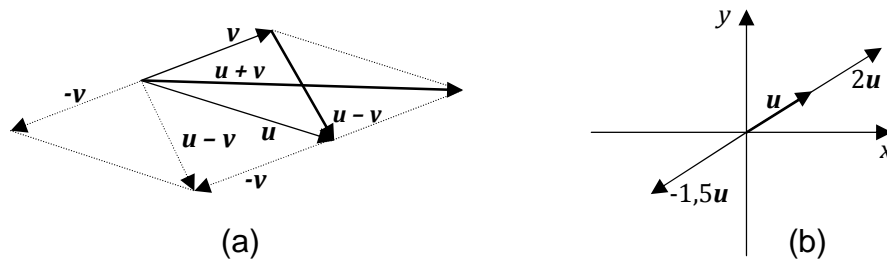


Figura 3 – Visão geométrica das operações com vetores: (a) soma e subtração de vetores; (b) multiplicação por escalar

Produto interno e norma

O produto interno entre os vetores u e v é dado por $u^t v = \langle u, v \rangle = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$.

A norma (ou comprimento) do vetor u , $\|u\|$, é dada por $\|u\| = \sqrt{u^t u}$, isto é, a raiz quadrada do produto interno de um vetor por ele mesmo. Assim, $\|u\|^2 = u^t u = x_1^2 + x_2^2 + \dots + x_n^2$. A Figura 4 apresenta a visão geométrica da norma de um vetor no \mathbb{R}^2 , sendo $x_1 = a$ e $x_2 = b$.

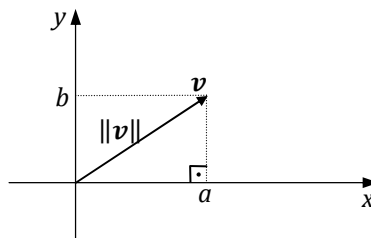


Figura 4 – Visão geométrica da norma de um vetor

Se um vetor v tem $\|v\| = 1$, então v é dito unitário. Chama-se de versor de um vetor v ao vetor unitário de mesma direção e sentido que v . Como consequência, $\text{versor}(v) = v' = \frac{v}{\|v\|}$. Diz-se que o versor de v é a forma normalizada do vetor v (Figura 5).

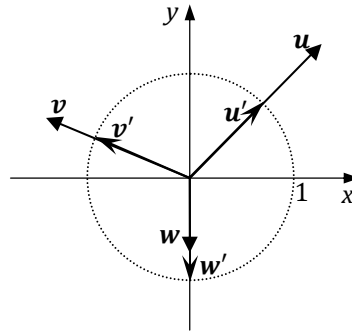


Figura 5 – Vetores no \mathbb{R}^2 e respectivos versores.

O produto interno entre dois vetores também pode ser dado pela fórmula $\mathbf{u}^t \mathbf{v} = \langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta$, sendo θ o ângulo entre os dois vetores. Assim, \mathbf{u} e \mathbf{v} são ortogonais, se e somente se, $\mathbf{u}^t \mathbf{v} = 0$.

A visão geométrica do produto interno é apresentada no item 2.1.2 – projeções ortogonais.

2.1.1.2 Matrizes

Uma matriz $A_{m \times n}$ é um arranjo retangular de elementos dispostos lado a lado em m linhas e n colunas, na qual o elemento a_{ij} representa o elemento da i -ésima linha com a j -ésima coluna:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix}$$

Uma matriz coluna é uma matriz $A_{n \times 1}$, também chamada de vetor. Uma matriz linha é uma matriz $A_{1 \times n}$, que pode ser identificada como a transposta de um vetor.

Uma matriz $A_{m \times n}$ pode ser vista como um arranjo de m vetores-linha com n coordenadas ou um arranjo de n vetores-coluna com m coordenadas.

$$A = \begin{bmatrix} \boxed{l_1} \\ \boxed{l_2} \\ \vdots \\ \boxed{l_m} \end{bmatrix} = \begin{bmatrix} \boxed{c_1} & \boxed{c_2} & \dots & \boxed{c_n} \end{bmatrix}$$

sendo $(l_i)_{1 \times n}$ e $(c_j)_{m \times 1}$.

Uma matriz A é dita quadrada se possui n linhas e n colunas.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

Uma matriz quadrada é dita diagonal se todos os elementos fora da diagonal principal são iguais à zero:

$$A = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{bmatrix} = \text{diag}(a_1, a_2, \dots, a_n)$$

A matriz identidade I é uma matriz quadrada de ordem n em que todos os elementos da diagonal principal são iguais a 1 e fora dela são iguais a 0:

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Operações com matrizes

São definidas as seguintes operações com matrizes:

(i) se A e B são matrizes de mesma ordem $m \times n$, então $A + B = C$, sendo C de ordem $m \times n$, tal que $c_{ij} = a_{ij} + b_{ij}$:

$$A + B = C = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \dots & a_{mn}+b_{mn} \end{bmatrix};$$

(ii) se A é uma matriz de ordem $m \times n$ e k é uma constante real, então a multiplicação por escalar é definida por $kA = C$, de mesma ordem que a matriz A , tal que $c_{ij} = ka_{ij}$:

$$C = kA = \begin{bmatrix} ka_{11} & ka_{12} & \dots & ka_{1n} \\ ka_{21} & ka_{22} & \dots & ka_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ka_{m1} & ka_{m2} & \dots & ka_{mn} \end{bmatrix};$$

(iii) se A é uma matriz de ordem $m \times n$ e B é uma matriz de ordem $n \times p$, então a multiplicação de A por B , $AB = C$ é uma matriz de ordem $m \times p$ tal que:

$$c_{ij} = \sum_{k=1}^n a_{ik} \cdot b_{kj}, \quad \text{com } i = 1, \dots, m \text{ e } j = 1, \dots, p.$$

Outra interpretação para o elemento c_{ij} é que ele representa o produto interno do vetor linha i da matriz A com o vetor coluna j da matriz B .

Tem-se que:

- (i) Normalmente $AB \neq BA$.
- (ii) A associação entre matrizes pode ser feita de várias formas, desde que se mantenha rigorosamente a ordem das matrizes envolvidas, como, por exemplo, em $(AB)(CD) = [A(BC)]D = A[(BC)D] = A[B(CD)]$.
- (iii) Se $AB = 0$ (matriz nula), não se pode afirmar que $A = 0$ ou $B = 0$.
- (iv) Se $AB = CB$, não se pode afirmar que $A = C$.

Se A é uma matriz de ordem $m \times n$, $B = \text{diag}(b_1, b_2, \dots, b_m)$ e $D = \text{diag}(d_1, d_2, \dots, d_n)$, então:

$$BA = \begin{bmatrix} \boxed{b_1 \mathbf{l}_1} \\ \boxed{b_2 \mathbf{l}_2} \\ \vdots \\ \boxed{b_m \mathbf{l}_m} \end{bmatrix} \text{ e } AD = \left[\boxed{d_1 \mathbf{c}_1} \quad \boxed{d_2 \mathbf{c}_2} \quad \dots \quad \boxed{d_n \mathbf{c}_n} \right].$$

Isto é, se uma matriz diagonal pré-multiplica uma matriz A , a matriz resultante é a matriz A tendo suas linhas multiplicadas pelos elementos correspondentes da matriz diagonal; e se uma matriz diagonal pós-multiplica

uma matriz A , a matriz resultante é a matriz A tendo suas colunas multiplicadas pelos elementos correspondentes da matriz diagonal.

Matriz transposta

Dada uma matriz A , sua transposta A^t é obtida com os mesmos elementos originais, mas escrevendo suas linhas como colunas e vice-versa. Assim, a transposta de

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \Rightarrow B = A^t = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}$$

Uma matriz quadrada A é dita simétrica quando $a_{ij} = a_{ji}$:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{12} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{bmatrix}$$

Numa matriz simétrica a linha k é idêntica à coluna k , o que faz com que a matriz seja igual à sua transposta, $A = A^t$.

São propriedades operatórias das matrizes simétricas:

- (i) $(A + B)^t = A^t + B^t$
- (ii) $(AB)^t = B^t A^t$
- (iii) $(A^t)^t = A$

Se A é uma matriz de ordem $m \times n$, então AA^t e $A^t A$ são matrizes quadradas e simétricas de ordem $m \times m$ e $n \times n$ respectivamente. Em ambos os casos, os elementos da diagonal principal são produtos internos de uma linha (ou coluna) por ela mesma, isto é, os elementos da diagonal principal representam o quadrado da norma do vetor linha (ou coluna) da matriz A :

$$AA^t = \begin{bmatrix} \langle \mathbf{l}_1, \mathbf{l}_1 \rangle & \langle \mathbf{l}_1, \mathbf{l}_2 \rangle & \dots & \langle \mathbf{l}_1, \mathbf{l}_m \rangle \\ \langle \mathbf{l}_2, \mathbf{l}_1 \rangle & \langle \mathbf{l}_2, \mathbf{l}_2 \rangle & \dots & \langle \mathbf{l}_2, \mathbf{l}_m \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{l}_m, \mathbf{l}_1 \rangle & \langle \mathbf{l}_m, \mathbf{l}_2 \rangle & \dots & \langle \mathbf{l}_m, \mathbf{l}_m \rangle \end{bmatrix} = \begin{bmatrix} \|\mathbf{l}_1\|^2 & \langle \mathbf{l}_1, \mathbf{l}_2 \rangle & \dots & \langle \mathbf{l}_1, \mathbf{l}_m \rangle \\ \langle \mathbf{l}_2, \mathbf{l}_1 \rangle & \|\mathbf{l}_2\|^2 & \dots & \langle \mathbf{l}_2, \mathbf{l}_m \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{l}_m, \mathbf{l}_1 \rangle & \langle \mathbf{l}_m, \mathbf{l}_2 \rangle & \dots & \|\mathbf{l}_m\|^2 \end{bmatrix}$$

$$\text{e } A^t A = \begin{bmatrix} \langle \mathbf{c}_1, \mathbf{c}_1 \rangle & \langle \mathbf{c}_1, \mathbf{c}_2 \rangle & \dots & \langle \mathbf{c}_1, \mathbf{c}_n \rangle \\ \langle \mathbf{c}_2, \mathbf{c}_1 \rangle & \langle \mathbf{c}_2, \mathbf{c}_2 \rangle & \dots & \langle \mathbf{c}_2, \mathbf{c}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{c}_n, \mathbf{c}_1 \rangle & \langle \mathbf{c}_n, \mathbf{c}_2 \rangle & \dots & \langle \mathbf{c}_n, \mathbf{c}_n \rangle \end{bmatrix} = \begin{bmatrix} \|\mathbf{c}_1\|^2 & \langle \mathbf{c}_1, \mathbf{c}_2 \rangle & \dots & \langle \mathbf{c}_1, \mathbf{c}_n \rangle \\ \langle \mathbf{c}_2, \mathbf{c}_1 \rangle & \|\mathbf{c}_2\|^2 & \dots & \langle \mathbf{c}_2, \mathbf{c}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{c}_n, \mathbf{c}_1 \rangle & \langle \mathbf{c}_n, \mathbf{c}_2 \rangle & \dots & \|\mathbf{c}_n\|^2 \end{bmatrix}$$

Determinante

O determinante de uma matriz quadrada A de ordem n é um número obtido a partir de operações específicas sobre seus elementos, sendo definido como:

$$\det A = \sum_{\rho} (-1)^J a_{1j_1} a_{2j_2} \dots a_{nj_n}$$

onde $J = J(j_1, \dots, j_n)$ é o número de inversões da permutação (j_1, \dots, j_n) das n colunas de A e ρ indica que a soma é estendida a todas as $n!$ permutações de $(1 \ 2 \ \dots \ n)$.

Como consequência da definição, o cálculo do determinante de uma matriz A , $\det A$, envolve:

- (i) o somatório de $n!$ produtos com n fatores cada um, sendo metade deles positivos e a outra metade negativos;
- (ii) cada produto envolve obrigatoriamente um único elemento de cada linha e um único de cada coluna;
- (iii) colocando os elementos de um produto em ordem crescente das linhas (colunas), o sinal desse produto é dado pelo número de inversões feitas das colunas (linhas) desses elementos, sendo positivo se o número de inversões for par e negativo caso contrário.
- (iv) reescrevendo a definição de determinante a partir da ordenação das colunas tem-se:

$$\det A = \sum_{\rho} (-1)^J a_{j_1 1} a_{j_2 2} \dots a_{j_n n}$$

Os determinantes possuem várias propriedades como consequência de sua definição e das observações acima. Para este trabalho, destacam-se:

- (i) $\det(AB) = \det(A) \cdot \det(B)$
- (ii) $\det(A^t) = \det(A)$
- (iii) $\det(\text{diag}(a_1, \dots, a_n)) = a_1 \cdot a_2 \cdot \dots \cdot a_n$

Uma matriz quadrada é dita singular se seu determinante é igual à zero.

Matriz inversa

Se A e B são matrizes quadradas de mesma ordem tais que $AB = I$, então B é dita inversa de A , denotada por $B = A^{-1}$, e A é dita inversa de B , denotada por $A = B^{-1}$.

São propriedades da inversa:

- (i) A inversa de uma matriz quadrada é única.
- (ii) A matriz A é inversível se, e somente se, $\det(A) \neq 0$, isto é, se A é uma matriz não singular.
- (iii) O determinante da matriz inversa A^{-1} é o inverso do determinante da matriz A , isto é, $\det(A^{-1}) = \frac{1}{\det(A)}$.
- (iv) $(A^t)^{-1} = (A^{-1})^t$, isto é, a inversa da transposta é a transposta da inversa.
- (v) $(A^{-1})^{-1} = A$
- (vi) $(AB)^{-1} = B^{-1}A^{-1}$
- (vii) Se A é uma matriz simétrica não singular, então sua inversa também é simétrica e não singular.
- (viii) Se $A = \text{diag}(a_1, a_2, \dots, a_n)$, então $A^{-1} = \text{diag}(1/a_1, 1/a_2, \dots, 1/a_n)$

Matriz ortogonal

Uma matriz quadrada inversível A é dita ortogonal se $A^{-1} = A^t$. Como, neste caso, $\det A = \det A^t = \det A^{-1}$, então o determinante de uma matriz ortogonal é igual a 1 ou igual a -1 .

Se A é ortogonal, então $AA^t = A^tA = I$. Isto quer dizer que

$$AA^t = \begin{bmatrix} \langle \mathbf{l}_1, \mathbf{l}_1 \rangle & \langle \mathbf{l}_1, \mathbf{l}_2 \rangle & \dots & \langle \mathbf{l}_1, \mathbf{l}_n \rangle \\ \langle \mathbf{l}_2, \mathbf{l}_1 \rangle & \langle \mathbf{l}_2, \mathbf{l}_2 \rangle & \dots & \langle \mathbf{l}_2, \mathbf{l}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{l}_n, \mathbf{l}_1 \rangle & \langle \mathbf{l}_n, \mathbf{l}_2 \rangle & \dots & \langle \mathbf{l}_n, \mathbf{l}_n \rangle \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

$$A^t A = \begin{bmatrix} \langle \mathbf{c}_1, \mathbf{c}_1 \rangle & \langle \mathbf{c}_1, \mathbf{c}_2 \rangle & \dots & \langle \mathbf{c}_1, \mathbf{c}_n \rangle \\ \langle \mathbf{c}_2, \mathbf{c}_1 \rangle & \langle \mathbf{c}_2, \mathbf{c}_2 \rangle & \dots & \langle \mathbf{c}_2, \mathbf{c}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{c}_n, \mathbf{c}_1 \rangle & \langle \mathbf{c}_n, \mathbf{c}_2 \rangle & \dots & \langle \mathbf{c}_n, \mathbf{c}_n \rangle \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Como o produto interno de duas linhas (ou colunas) distintas tem que ser zero, então as linhas (ou colunas) de A são vetores ortogonais entre si. Como o produto interno de uma linha (ou coluna) consigo mesma tem que ser igual a 1, então $\|\mathbf{l}_i\|^2 = 1$ ($\|\mathbf{c}_i\|^2 = 1$), isto é, a norma de cada vetor linha (ou vetor coluna) de A é sempre igual a 1.

Posto e traço

Dada uma matriz A de ordem $m \times n$, uma submatriz de A é qualquer matriz obtida de A pela supressão de alguma(s) linha(s) ou coluna(s).

O posto ou *rank* de uma matriz A , $p(A)$ ou $r(A)$ ou *posto* (A) ou *rank*(A), é dado pela ordem da maior submatriz quadrada de A que possui determinante diferente de zero. O posto de uma matriz A pode ser obtido, também, pelo número de linhas ou colunas linearmente independentes que A possui, como poderá ser visto adiante. Como consequência, $p(A) \leq \min\{m, n\}$.

O traço de uma matriz quadrada de ordem n , $tr(A)$ ou *traço* (A), é dado pela soma dos elementos da diagonal principal da matriz, isto é, $tr(A) = \sum_{i=1}^n a_{ii}$.

2.1.1.3 Espaço vetorial

De uma maneira geral, espaço vetorial é todo conjunto não vazio cujos elementos se comportam como vetores, isto é, seus elementos operam como vetores e mantêm as propriedades dos vetores. Assim, os espaços vetoriais possuem as seguintes propriedades quando \mathbf{u} , \mathbf{v} e \mathbf{w} são elementos do espaço vetorial V e a e b são constantes reais:

- (i) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

- (ii) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
- (iii) $\mathbf{u} + \mathbf{0} = \mathbf{u}, \forall \mathbf{u} \in V$
- (iv) $\forall \mathbf{u} \in V, \exists (-\mathbf{u}) \in V: \mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
- (v) $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$
- (vi) $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$
- (vii) $a(b\mathbf{u}) = (ab)\mathbf{u}$
- (viii) $1\mathbf{u} = \mathbf{u}$

Os espaços vetoriais que serão usados neste trabalho serão \mathbb{R} (unidimensional), \mathbb{R}^2 (bidimensional), \mathbb{R}^3 (tridimensional), ..., \mathbb{R}^n (n-dimensional).

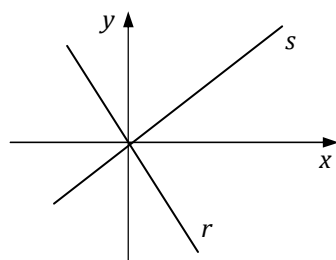
Subespaço vetorial

Dado um espaço vetorial V , um subconjunto W de V não vazio, será um subespaço vetorial de V se:

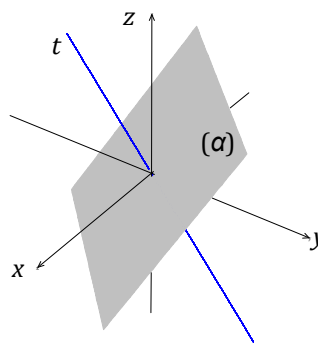
- (i) $\forall \mathbf{u}, \mathbf{v} \in W$, tivermos $(\mathbf{u} + \mathbf{v}) \in W$
- (ii) $\forall a \in \mathbb{R}, \forall \mathbf{u} \in W$, tivermos $(a\mathbf{u}) \in W$

Estas duas condições garantem que W tenha as 8 propriedades de espaço vetorial e que um subespaço vetorial sempre vai conter o vetor nulo $\mathbf{0}$.

Qualquer reta que passe pela origem do sistema de eixos do \mathbb{R}^n é um subespaço desse espaço vetorial (Figura 6), assim como qualquer plano contendo a origem também é subespaço do \mathbb{R}^n , com $n \geq 2$.



(a)



(b)

Figura 6 – Exemplos de subespaços (a) do \mathbb{R}^2 : s e r ; (b) do \mathbb{R}^3 : t e α

Dependência e independência linear

Se um vetor v é combinação linear de u_1, u_2, \dots, u_n , então v pode ser obtido a partir desses vetores, de tal forma que $v = a_1 u_1 + a_2 u_2 + \dots + a_n u_n$ conforme ilustrado na Figura 7.

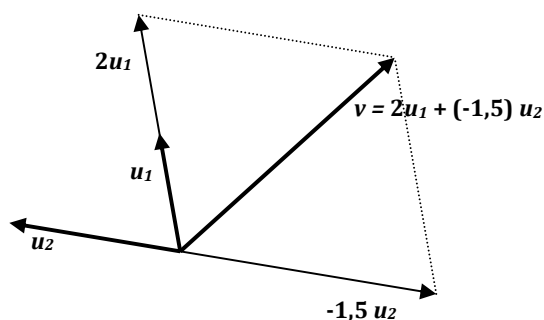


Figura 7 – Visão geométrica de um exemplo de combinação linear

Se em u_1, u_2, \dots, u_n existir algum vetor que é combinação linear dos outros, então diz-se que esses vetores são linearmente dependentes, e, caso contrário, os vetores são ditos linearmente independentes.

Base de um espaço vetorial

Seja V um espaço vetorial. Qualquer subconjunto não vazio de k vetores de V , $\{u_1, u_2, \dots, u_k\}$, é capaz de gerar um subespaço vetorial W , $[u_1, u_2, \dots, u_k] = W$, contido em V .

Seja $\beta = \{u_1, u_2, \dots, u_n\}$ o menor subconjunto de vetores linearmente independentes capaz de gerar W . Neste caso diz-se que β é uma base de W e a dimensão de W é dada pelo número de vetores da base, $\dim(W) = n$. Então, uma base de um espaço vetorial W é um conjunto de vetores linearmente independentes capaz de gerar esse espaço vetorial. Um mesmo espaço vetorial pode apresentar muitas bases, tendo todas o mesmo número de elementos.

Seja V um espaço vetorial com base $\beta = \{u_1, u_2, \dots, u_n\}$. Se v é um vetor qualquer desse espaço vetorial, então v pode ser escrito como combinação linear dos vetores da base β , ou seja, $v = a_1 u_1 + a_2 u_2 + \dots + a_n u_n$. Assim, define-se o vetor de coordenadas de v na base β por:

$$[v]_{\beta} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

Se B a matriz cujos vetores coluna são os vetores da base β , então $B \cdot [v]_{\beta} = v$.

Se todos os vetores de uma base são unitários, então essa base é chamada de base normal; se todos os vetores de uma base são ortogonais entre si, então essa base é chamada de base ortogonal; se uma base for normal e ortogonal, então ela é chamada de base ortonormal.

Seja $\beta = \{u_1, u_2, \dots, u_n\}$ uma base ortogonal para o espaço vetorial V e B a matriz de vetores coluna dessa base. Então,

$$B^t B = \begin{bmatrix} \langle u_1, u_1 \rangle & \langle u_1, u_2 \rangle & \dots & \langle u_1, u_n \rangle \\ \langle u_2, u_1 \rangle & \langle u_2, u_2 \rangle & \dots & \langle u_2, u_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle u_n, u_1 \rangle & \langle u_n, u_2 \rangle & \dots & \langle u_n, u_n \rangle \end{bmatrix} = \begin{bmatrix} \|c_1\|^2 & 0 & \dots & 0 \\ 0 & \|c_2\|^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \|c_n\|^2 \end{bmatrix}$$

Mas, se β for ortonormal, então $B^t B = B B^t = I$. Logo, os vetores linha da matriz B também formam uma base ortonormal do espaço vetorial V .

2.1.1.4 Sistemas lineares

Um sistema de equações lineares é um conjunto de equações

$$S: \begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{cases}$$

que pode ser representado matricialmente por

$$Ax = b$$

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \text{ sendo}$$

A a matriz de coeficientes, x o vetor composto de n variáveis e b o vetor de termos independentes.

Uma solução de um sistema linear é um vetor que satisfaz todas as equações do sistema simultaneamente. Um sistema linear pode ter ou não solução.

Se o sistema possui uma única solução, então ele é dito sistema possível determinado (SPD) e $\text{posto}(A) = \text{posto}(A|b) = n$, sendo $A|b$ a matriz A ampliada pela justaposição do vetor coluna b . Se A é uma matriz quadrada e $\det(A) \neq 0$ (A é não singular) então a solução única do sistema é dada por $x = A^{-1}b$.

Se o sistema possui infinitas soluções, então ele é dito sistema possível indeterminado (SPI) e $\text{posto}(A) = \text{posto}(A|b) < n$.

Se o sistema não possui solução, então ele é dito sistema impossível (SI) e $\text{posto}(A) \neq \text{posto}(A|b)$.

Um sistema é dito homogêneo se o vetor de termos independentes é nulo:

$$Ax = \mathbf{0}$$

$$S: \begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = 0 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = 0 \end{cases}$$

Todo sistema homogêneo é possível, pois possui a solução trivial $[x_1 \ x_2 \ \dots \ x_n]^t = [0 \ 0 \ \dots \ 0]^t$. Se o sistema é SPD, esta solução é única. Se o sistema é SPI, há ainda infinitas soluções para o sistema. Se a matriz A do sistema é quadrada e singular, então o sistema será obrigatoriamente SPI.

Para determinar a solução de um sistema linear um dos processos mais utilizados é o método de escalonamento que é baseado na transformação da matriz ampliada $A|b$ numa matriz escada. Para maiores esclarecimentos consultar a bibliografia dada.

2.1.2 Projeções ortogonais

Sejam v_1 e v_2 dois vetores do \mathbb{R}^2 , tais que $v_1^t v_2 = v_2^t v_1 = 0$, então v_1 e v_2 são ortogonais entre si. Tais vetores formam uma base ortogonal do \mathbb{R}^2 , isto

é, qualquer outro vetor do \mathbb{R}^2 pode ser escrito como combinação linear dos vetores da base $\{v_1, v_2\}$, conforme exemplificado a seguir.

Sejam os vetores w, v_1 e v_2 representados no plano xy através de suas respectivas coordenadas tais como $w = (2,2), v_1 = (3,-1), v_2 = (1,3)$. Como v_1 e v_2 são ortogonais entre si, pode-se utilizá-los como uma base para reescrever as coordenadas de w através de uma combinação linear entre esses vetores, em que $w = c_1 \cdot v_1 + c_2 \cdot v_2$, cuja solução é trivial e dada por:

$$(2, 2) = \frac{2}{5} \cdot v_1 + \frac{4}{5} \cdot v_2$$

Graficamente, tem-se a seguinte representação:

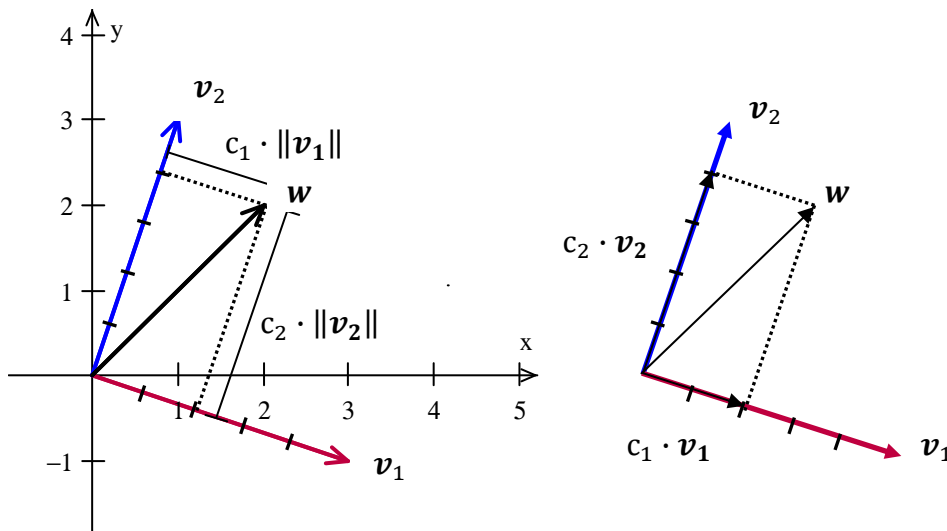


Figura 8 – Gráfico comparativo ilustrando a mudança de base do vetor w .

Chamam-se de projeções ortogonais de w sobre os vetores da base v_1 e v_2 , respectivamente aos vetores $(Proj w)_{v_1} = \frac{2}{5}v_1$ e $(Proj w)_{v_2} = \frac{4}{5}v_2$.

Graficamente tem-se:

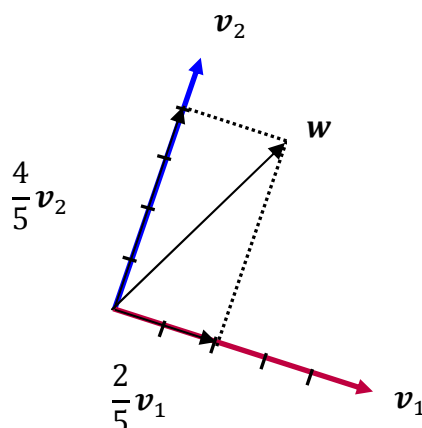


Figura 9 – Projeções ortogonais de w sobre os vetores da base v_1 e v_2 .

Os conceitos de base ortogonal e projeções ortogonais sobre os vetores da base podem ser estendidos para o \mathbb{R}^n . Portanto, se $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ são n vetores do \mathbb{R}^n ortogonais entre si, então esses vetores formam uma base ortogonal do \mathbb{R}^n . Se, além disso, esses n vetores são unitários, isto é, $\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \dots = \|\mathbf{v}_n\| = 1$, então a base formada é ortonormal.

Assim, dado um vetor \mathbf{w} do \mathbb{R}^n , esse vetor pode ser escrito como uma combinação linear dos vetores da base ortonormal $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$:

$$\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n$$

Tem-se que as projeções ortogonais de \mathbf{w} sobre os vetores da base ortonormal são dadas por:

$$\begin{aligned} (Proj \mathbf{w})_{\mathbf{v}_1} &= c_1\mathbf{v}_1 \\ (Proj \mathbf{w})_{\mathbf{v}_2} &= c_2\mathbf{v}_2 \\ &\vdots \\ (Proj \mathbf{w})_{\mathbf{v}_n} &= c_n\mathbf{v}_n \end{aligned}$$

Portanto, para que cada projeção seja definida, é preciso determinar c_1, c_2, \dots, c_n . Quando a base é ortonormal estes cálculos são tremendamente facilitados e obtém-se $c_i = \mathbf{w}^t \mathbf{v}_i$ conforme ilustrado a seguir.

$$\begin{aligned} \mathbf{w} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n \\ \mathbf{w}^t &= (c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n)^t = c_1\mathbf{v}_1^t + c_2\mathbf{v}_2^t + \dots + c_n\mathbf{v}_n^t \end{aligned}$$

Pós multiplicando-se por \mathbf{v}_1 ambos os lados da igualdade, tem-se:

$$\begin{aligned} \mathbf{w}^t \mathbf{v}_1 &= (c_1\mathbf{v}_1^t + c_2\mathbf{v}_2^t + \dots + c_n\mathbf{v}_n^t)\mathbf{v}_1 \\ \mathbf{w}^t \mathbf{v}_1 &= c_1\mathbf{v}_1^t \mathbf{v}_1 + c_2\mathbf{v}_2^t \mathbf{v}_1 + \dots + c_n\mathbf{v}_n^t \mathbf{v}_1 \end{aligned}$$

como $\mathbf{v}_i^t \mathbf{v}_j = 0$, se $i \neq j$ (vetores ortogonais) e $\mathbf{v}_i^t \mathbf{v}_i = \|\mathbf{v}_i\|^2 = 1$

$$\mathbf{w}^t \mathbf{v}_1 = c_1 1 + c_2 0 + \dots + c_n 0 = c_1$$

Analogamente, todo c_i é dado pelo produto interno entre \mathbf{w} e \mathbf{v}_i ,

$$c_i = \mathbf{w}^t \mathbf{v}_i = \mathbf{v}_i^t \mathbf{w}$$

Assim:

$$(\text{Proj } \mathbf{w})_{\mathbf{v}_i} = c_i \mathbf{v}_i = (\mathbf{w}^t \mathbf{v}_i) \mathbf{v}_i = (\mathbf{v}_i^t \mathbf{w}) \mathbf{v}_i \quad (1)$$

O resultado acima permite a visualização geométrica do produto interno entre dois vetores: o produto interno entre os vetores \mathbf{u} e \mathbf{v} é a norma da projeção ortogonal de \mathbf{u} sobre \mathbf{v} , se \mathbf{v} é um vetor unitário (Figura 10).

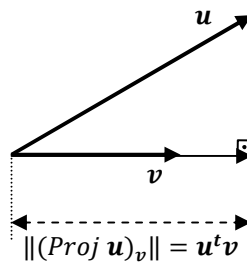


Figura 10 – Visão geométrica do produto interno entre os vetores \mathbf{u} e \mathbf{v} , tendo \mathbf{v} norma 1

Se a norma de \mathbf{v} é maior do que zero e diferente de 1, então

$$\|(\text{Proj } \mathbf{u})_{\mathbf{v}}\| = \mathbf{u}^t \frac{\mathbf{v}}{\|\mathbf{v}\|}.$$

2.1.3 Autovalores e autovetores

O estudo de autovalores ou raízes características e os seus respectivos autovetores são de fundamental importância na diagonalização e decomposição de uma matriz de ordem $n \times n$, visto que tal procedimento facilita o entendimento global da matriz em relação ao seu determinante, inversa se existir, cálculo de potência de matrizes e projeções ortogonais que podem ser obtidas através da decomposição espectral. Outras generalizações podem ser feitas para matrizes de ordem $m \times n$, como a decomposição em valores singulares que são utilizadas na AC.

Para facilitar o entendimento, o estudo de autovalores e autovetores será abordado na perspectiva de transformações lineares.

2.1.3.1 Transformações lineares

Definição: sejam \mathbb{R}^n e \mathbb{R}^m dois espaços vetoriais de dimensões n e m respectivamente, em que n e m podem ou não serem diferentes. Uma transformação linear é uma função do \mathbb{R}^n no \mathbb{R}^m , isto é, $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$, que satisfaz as seguintes condições (Poole, 2004, p. 188):

a) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$, para todo \mathbf{u} e \mathbf{v} em \mathbb{R}^n .

b) $T(c\mathbf{v}) = cT(\mathbf{v})$, para todo \mathbf{v} em \mathbb{R}^n e todos os escalares c em \mathbb{R} .

Ambas as condições acima significam que uma transformação linear preserva a soma entre vetores e a multiplicação de vetor por escalar e, como consequência do item b, a transformação do vetor nulo leva no vetor nulo, isto é, $T(\mathbf{0}) = \mathbf{0}$.

Considera-se que existe uma transformação linear quando um vetor é modificado seguindo-se algumas regras e preservando-se algumas propriedades, tais como soma e multiplicação por escalar. Por exemplo, seja T uma transformação linear de um vetor \mathbf{v} em \mathbf{w} pertencentes aos espaços vetoriais \mathbb{R}^3 e \mathbb{R}^2 respectivamente, tais que $\mathbf{v} = (x, y, z)$ e $\mathbf{w} = (x + y, -x + 2z)$. Abaixo descreve-se tal transformação:

$$T: \mathbb{R}^3 \rightarrow \mathbb{R}^2$$
$$T(x, y, z) = (x + y, -x + 2z)$$

Esta transformação pode ser facilmente estabelecida como $A\mathbf{v} = \mathbf{w}$ com a seguinte operação matricial

$$\begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x + y \\ -x + 2z \end{bmatrix} \text{ em que:}$$
$$A = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 2 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \text{ e} \quad \mathbf{w} = \begin{bmatrix} x + y \\ -x + 2z \end{bmatrix}$$

Assim, o vetor \mathbf{v} é transformado em \mathbf{w} pela matriz de transformação A , que dita às regras da transformação, ou seja, a matriz de transformação A estabelece o relacionamento entre os dois vetores \mathbf{v} e \mathbf{w} .

Portanto, uma forma geral para uma transformação linear qualquer do \mathbb{R}^n no \mathbb{R}^m é descrita como:

$$\begin{aligned} T: \mathbb{R}^n &\rightarrow \mathbb{R}^m \\ T(\mathbf{v}) &= A\mathbf{v} = \mathbf{w} \end{aligned} \quad (2)$$

em que A é uma matriz de ordem $m \times n$, sendo $\mathbf{v} \in \mathbb{R}^n$ e $\mathbf{w} \in \mathbb{R}^m$.

Quando uma transformação linear ocorre dentro do próprio espaço vetorial, isto é, $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$, esta é chamada de operador linear.

2.1.3.2 Relação entre transformações lineares e autovalores e autovetores

Em alguns operadores lineares é de interesse que a transformação de \mathbf{v} pela matriz A leve a um vetor \mathbf{w} que é um múltiplo escalar de \mathbf{v} , de modo que ele possa ser reescrito em função de \mathbf{v} , conforme descrito abaixo:

$$T(\mathbf{v}) = A\mathbf{v} = \mathbf{w} = \lambda\mathbf{v} \quad (3)$$

Neste caso, λ é o *autovalor* de A e \mathbf{v} é o *autovetor* associado ao autovalor λ , sendo $\mathbf{v} \neq \vec{0}$

A figura a seguir ilustra a comparação de uma transformação linear com e sem geração de um autovetor. Observe que em (a) a transformação do vetor \mathbf{v} preserva a direção de \mathbf{v} , isto é, \mathbf{v} e $\mathbf{w} = A\mathbf{v}$ estão na mesma direção, o que caracteriza que \mathbf{v} é um autovetor associado a um $\lambda = 3$. Em (b) \mathbf{v} e $\mathbf{w} = A\mathbf{v}$ não estão na mesma direção, logo \mathbf{v} não é um autovetor.

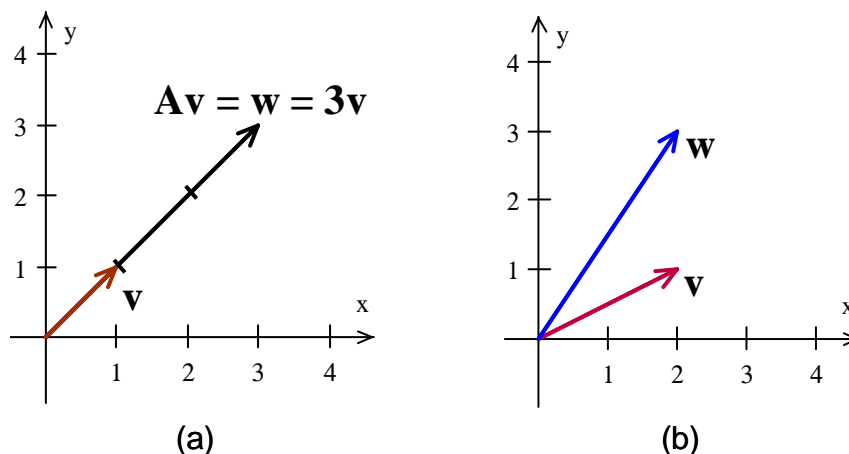


Figura 11 – Comparação da transformação linear $T(\mathbf{v}) = A\mathbf{v} = \mathbf{w}$ com geração de autovetor (a) e sem a geração do mesmo (b)

O exemplo a seguir obtido de Lay (1999, p. 273) ilustra a utilização de uma mesma matriz A para realizar duas transformações lineares dos vetores \mathbf{u} e \mathbf{v} em $A\mathbf{u}$ e $A\mathbf{v}$, sendo que somente a transformação $A\mathbf{v}$ é um múltiplo escalar de \mathbf{v} , isto é, $A\mathbf{v} = \lambda\mathbf{v}$ em que $\lambda = -4$ é um autovalor de A . Já a transformação $A\mathbf{u}$ não é um múltiplo escalar de \mathbf{u} e portanto \mathbf{u} não é um autovetor de A .

$$A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 6 \\ -5 \end{bmatrix} \quad \text{e} \quad \mathbf{u} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

$$A\mathbf{v} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix}, \text{ portanto}$$

$$A\mathbf{v} = -4 \begin{bmatrix} 6 \\ -5 \end{bmatrix} = \lambda\mathbf{v}$$

$$A\mathbf{u} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 11 \end{bmatrix}, \text{ portanto}$$

$$A\mathbf{u} \neq \lambda \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

Um importante teorema abordado a seguir garante a existência de infinitos autovetores associados ao mesmo autovalor λ .

Teorema: Dado um operador linear $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ e um autovetor \mathbf{v} associado ao autovalor λ , qualquer vetor $\mathbf{u} = c\mathbf{v}$ ($c \in \mathbb{R}$ e $c \neq 0$) é também autovetor de T associado a λ .

Demonstração:

Seja \mathbf{v} um autovetor associado a λ , portanto $T(\mathbf{v}) = \lambda\mathbf{v}$, e considere um vetor \mathbf{u} proporcional a \mathbf{v} de modo que $\mathbf{u} = c\mathbf{v}$. Tem-se que:

$$T(\mathbf{u}) = T(c\mathbf{v}) = cT(\mathbf{v}) = c\lambda\mathbf{v} = \lambda c\mathbf{v} = \lambda\mathbf{u}$$

Logo, $T(\mathbf{u}) = \lambda\mathbf{u}$, então \mathbf{u} também é um autovetor associado ao autovalor λ . Portanto, qualquer múltiplo escalar de um autovetor é também um autovetor associado ao mesmo autovalor.

Pode-se visualizar o teorema anterior utilizando-se o exemplo de Poole (2004, p. 233) descrito a seguir:

$$\text{Dado } A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ e } \lambda = 4, \text{ portanto } A\mathbf{u} = \lambda\mathbf{u}.$$

$$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Então \mathbf{u} é um autovetor associado a λ . Note que os vetores $\mathbf{w} = 2\mathbf{u}$ e $\mathbf{t} = 4\mathbf{u}$ descritos abaixo também são autovetores associados ao autovalor $\lambda = 4$.

$$\mathbf{w} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \text{e} \quad \mathbf{t} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}, \text{ portanto,}$$

$$\begin{array}{ll} A\mathbf{w} = \lambda\mathbf{w} & A\mathbf{t} = \lambda\mathbf{t} \\ \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = 4 \begin{bmatrix} 2 \\ 2 \end{bmatrix} & \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 4 \end{bmatrix} = 4 \begin{bmatrix} 4 \\ 4 \end{bmatrix} \end{array}$$

Cálculo dos autovalores e autovetores

Existem diversos procedimentos para facilitar o cálculo dos autovalores e autovetores. O procedimento mostrado a seguir apresenta uma fundamentação simples e segura.

Sendo \mathbf{v} um autovetor de A associado a um autovalor λ , então:

$$\begin{aligned} A\mathbf{v} &= \lambda\mathbf{v}, & \text{para } \mathbf{v} \neq \mathbf{0} \\ A\mathbf{v} - \lambda\mathbf{v} &= \mathbf{0}, \\ A\mathbf{v} - \lambda I\mathbf{v} &= \mathbf{0}, \\ (A - \lambda I)\mathbf{v} &= \mathbf{0}. \end{aligned}$$

Desta forma, encontrar os autovalores de A equivale a resolver o sistema homogêneo $(A - \lambda I)\mathbf{v} = \mathbf{0}$, isto é, encontrar os vetores $\mathbf{v} \neq \mathbf{0}$, que satisfaçam essa igualdade. Um sistema é homogêneo quando todos os termos independentes (lado direito da igualdade) são iguais à zero. Obviamente, todo sistema homogêneo admite a solução trivial, que é o próprio vetor nulo. Como \mathbf{v} deve ser diferente do vetor nulo, o sistema homogêneo $(A - \lambda I)\mathbf{v} = \mathbf{0}$ deverá ter solução além da trivial. Portanto, o sistema precisa ser possível e indeterminado, isto é, ele deverá ter infinitas soluções. Para isso, $(A - \lambda I)$ deverá ser singular, ou seja, $\det(A - \lambda I) = 0$.

Sendo:

$$A - \lambda I = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$A - \lambda I = \begin{bmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{bmatrix}$$

Então, $P(\lambda) = \det(A - \lambda I)$ é um polinômio em λ de grau n , chamado de polinômio característico da matriz A . A equação $\det(A - \lambda I) = 0$ é chamada de equação característica de A .

Assim, pode-se determinar os autovalores de A encontrando-se as raízes do polinômio característico, o qual terá no máximo n raízes reais, em que n é a ordem da matriz. A seguir, um exemplo ilustrativo, apresentado por Leon (1999, p. 214) para o qual são calculados (com detalhes algébricos) os autovalores associados a uma matriz A e os respectivos autovetores.

Seja $A = \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix}$, portanto o sistema a ser resolvido $(A - \lambda I)v = 0$,

para o qual se tem,

$$A - \lambda I = \begin{bmatrix} 2 - \lambda & -3 & 1 \\ 1 & -2 - \lambda & 1 \\ 1 & -3 & 2 - \lambda \end{bmatrix}.$$

O resultado desejado obtido de $\det(A - \lambda I) = 0$, fornece,

$$(2 - \lambda) \det \begin{bmatrix} -2 - \lambda & 1 \\ -3 & 2 - \lambda \end{bmatrix} - (-3) \det \begin{bmatrix} 1 & 1 \\ 1 & 2 - \lambda \end{bmatrix} + 1 \det \begin{bmatrix} 1 & -2 - \lambda \\ 1 & -3 \end{bmatrix} = 0$$

$$(2 - \lambda)[(-2 - \lambda)(2 - \lambda) + 3] + 3[(2 - \lambda) - 1] + [(-3) - (-2 - \lambda)] = 0$$

$$(2 - \lambda)^2(-2 - \lambda) + 3(2 - \lambda) + 3 - 3\lambda - 1 + \lambda = 0$$

$$(4 - 4\lambda + \lambda^2)(-2 - \lambda) + 6 - 3\lambda + 2 - 2\lambda = 0$$

$$-8 - 4\lambda + 8\lambda + 4\lambda^2 - 2\lambda^2 - \lambda^3 + 8 - 5\lambda = 0$$

$$-\lambda^3 + 2\lambda^2 - \lambda = 0 \quad \cdot (-1)$$

$$\lambda^3 - 2\lambda^2 + \lambda = 0$$

$$\lambda(\lambda^2 - 2\lambda + 1) = 0$$

Tem-se que $\lambda_1 = 0$ ou $\lambda^2 - 2\lambda + 1 = 0$

$$\lambda^2 - 2\lambda + 1 = 0$$

$$\Delta = (-2)^2 - 4$$

$$\Delta = 0$$

$$\lambda_2 = \frac{-(-2) + \sqrt{0}}{2} = 1$$

$$\lambda_3 = \frac{-(-2) - \sqrt{0}}{2} = 1$$

Portanto as raízes da equação característica ou autovalores de A são:

$$\lambda_1 = 0 \quad \lambda_2 = \lambda_3 = 1$$

Para determinar os autovetores associados a cada autovalor, deve-se resolver o seguinte sistema:

$$(A - \lambda I)v = 0$$

Para $\lambda_1 = 0$, em que $v = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$, tem-se:

$$A - 0I = \begin{bmatrix} 2-0 & -3 & 1 \\ 1 & -2-0 & 1 \\ 1 & -3 & 2-0 \end{bmatrix} = \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix}$$

$$(A - 0I)v = 0$$

$$\begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Escalonando, tem-se:

$$\begin{bmatrix} 2 & -3 & 1 & : & 0 \\ 1 & -2 & 1 & : & 0 \\ 1 & -3 & 2 & : & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 1 & : & 0 \\ 2 & -3 & 1 & : & 0 \\ 1 & -3 & 2 & : & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 1 & : & 0 \\ 0 & 1 & -1 & : & 0 \\ 1 & -3 & 2 & : & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -2 & 1 & : & 0 \\ 0 & 1 & -1 & : & 0 \\ 0 & -1 & 1 & : & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & 1 & : & 0 \\ 0 & 1 & -1 & : & 0 \\ 0 & 0 & 0 & : & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 & : & 0 \\ 0 & 1 & -1 & : & 0 \\ 0 & 0 & 0 & : & 0 \end{bmatrix}$$

Então:

$$x - z = 0 \quad \therefore \quad x = z$$

$$y - z = 0 \quad \therefore \quad y = z$$

$$x = y = z \quad \therefore$$

Assim,

$$\mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} z \\ z \\ z \end{bmatrix} = z \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \Rightarrow \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Para $\lambda_2 = \lambda_3 = 1$, em que $\mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$, tem-se:

$$A - 1I = \begin{bmatrix} 2-1 & -3 & 1 \\ 1 & -2-1 & 1 \\ 1 & -3 & 2-1 \end{bmatrix} = \begin{bmatrix} 1 & -3 & 1 \\ 1 & -3 & 1 \\ 1 & -3 & 1 \end{bmatrix}$$
$$(A - 1I)\mathbf{v} = 0$$

Escalonando, tem-se:

$$\left[\begin{array}{ccc|c} 1 & -3 & 1 & 0 \\ 1 & -3 & 1 & 0 \\ 1 & -3 & 1 & 0 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & -3 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -3 & 1 & 0 \end{array} \right] \sim \left[\begin{array}{ccc|c} 1 & -3 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

Então:

$$x - 3y + z = 0 \quad \therefore$$

$$x = 3y - z \quad \therefore$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3y - z \\ y \\ z \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3y \\ y \\ 0 \end{bmatrix} + \begin{bmatrix} -z \\ 0 \\ z \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = y \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

$\mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$ e $\mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ é uma das soluções possíveis para os vetores \mathbf{v}_2

e \mathbf{v}_3 assumir associado a $\lambda_2 = \lambda_3 = 1$.

2.1.4 Diagonalização e decomposição de matrizes

Até este ponto no texto foram apresentados alguns conceitos importantes relacionados à AC, porém não diretamente envolvidos nos cálculos desta metodologia e sim para um melhor entendimento da próxima seção, que trata de métodos para diagonalização e decomposição de matrizes, os quais são cruciais para o entendimento da metodologia AC. Em especial a decomposição em valores singulares (DVS) que é a base da AC, entretanto, antes de abordar diretamente a DVS, é necessário que se apresente as decomposições para matrizes quadradas e simétricas.

2.1.4.1 Diagonalização e decomposição de matrizes quadradas

Mostra-se a seguir como utilizar os n autovalores e respectivos autovetores para diagonalizar ou decompor uma matriz A , quadrada de ordem n .

Seja D a matriz diagonal contendo os autovalores λ_i de A , isto é $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, sendo $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

Seja P uma matriz cujas colunas são os autovetores de A , de modo que o autovetor \mathbf{v}_i está associado ao autovalor λ_i ,

$$P = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n] = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix}.$$

Então $A\mathbf{v}_i = \lambda_i\mathbf{v}_i = \mathbf{w}_i$, para $i = 1, 2, \dots, n$, definem as transformações dos autovetores de A . Seja W uma matriz cujas colunas são estas transformações:

$$\begin{aligned}
W &= [\mathbf{w}_1 \quad \mathbf{w}_2 \quad \cdots \quad \mathbf{w}_n] \\
W &= [A\mathbf{v}_1 \quad A\mathbf{v}_2 \quad \cdots \quad A\mathbf{v}_n] \\
W &= [\lambda_1\mathbf{v}_1 \quad \lambda_2\mathbf{v}_2 \quad \cdots \quad \lambda_n\mathbf{v}_n].
\end{aligned}$$

Portanto,

$$[A\mathbf{v}_1 \quad A\mathbf{v}_2 \quad \cdots \quad A\mathbf{v}_n] = [\lambda_1\mathbf{v}_1 \quad \lambda_2\mathbf{v}_2 \quad \cdots \quad \lambda_n\mathbf{v}_n],$$

de modo que,

$$\begin{aligned}
A[\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n] &= \begin{bmatrix} \lambda_1 x_{11} & \lambda_2 x_{12} & \cdots & \lambda_n x_{1n} \\ \lambda_1 x_{21} & \lambda_2 x_{22} & \cdots & \lambda_n x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 x_{n1} & \lambda_2 x_{n2} & \cdots & \lambda_n x_{nn} \end{bmatrix} \\
AP &= \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = PD.
\end{aligned}$$

Assim:

$$AP = PD$$

A igualdade $AP = PD$ quando pré-multiplicada por P^{-1} fornece a diagonalização da matriz A , e quando pós-multiplicada por P^{-1} fornece a decomposição da matriz A . Então, obviamente, é necessário que exista a inversa P^{-1} da matriz P .

Pré-multiplicando,

$$P^{-1}AP = P^{-1}PD = D,$$

$$P^{-1}AP = D.$$

Pós-multiplicando,

$$APP^{-1} = A = PDP^{-1}$$

$$A = PDP^{-1}.$$

Então, se existir uma matriz P inversível tal que $P^{-1}AP = D$, diz-se que A é uma matriz diagonalizável e semelhante a D . Já a expressão $A = PDP^{-1}$ representa uma decomposição da matriz A em autovalores e autovetores.

Se a matriz A é semelhante à matriz diagonal D , denota-se $A \sim D$, as seguintes condições podem ser verificadas:

- $\det A = \det D$;
- A é inversível, se e somente se, D é inversível;
- A e D têm o mesmo posto;

- d) A e D têm o mesmo polinômio característico e por consequência, os mesmo autovalores.

2.1.4.2 Diagonalização e decomposição de matrizes simétricas

Um importante teorema da álgebra linear diz que toda matriz A simétrica é diagonalizável. Além disso, seus autovetores associados a autovalores distintos são ortogonais entre si.

Para os autovetores associados a autovalores iguais, é possível ortogonalizá-los, sendo o algoritmo de Gram-Schmidt uma das alternativas mais comumente utilizadas.

Para uma matriz A simétrica é possível obter uma matriz P de autovetores com todos eles ortogonais entre si.

Um fato interessante associado a essa matriz P está esquematicamente mostrado a seguir:

$$\mathbf{v}_1 = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix}; \quad \mathbf{v}_2 = \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{bmatrix}; \quad \dots; \quad \mathbf{v}_n = \begin{bmatrix} x_{1n} \\ x_{2n} \\ \vdots \\ x_{nn} \end{bmatrix}$$

$$P = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n] = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix}$$

$$P^t P = \begin{bmatrix} \mathbf{v}_1^t \\ \mathbf{v}_2^t \\ \vdots \\ \mathbf{v}_n^t \end{bmatrix} [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_n]$$

$$P^t P = \begin{bmatrix} \mathbf{v}_1^t \mathbf{v}_1 & \mathbf{v}_1^t \mathbf{v}_2 & \dots & \mathbf{v}_1^t \mathbf{v}_n \\ \mathbf{v}_2^t \mathbf{v}_1 & \mathbf{v}_2^t \mathbf{v}_2 & \dots & \mathbf{v}_2^t \mathbf{v}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_n^t \mathbf{v}_1 & \mathbf{v}_n^t \mathbf{v}_2 & \dots & \mathbf{v}_n^t \mathbf{v}_n \end{bmatrix}$$

Como os vetores $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ são ortogonais entre si, então $\mathbf{v}_i^t \mathbf{v}_j = 0$, se $i \neq j$ e $\mathbf{v}_i^t \mathbf{v}_i = \|\mathbf{v}_i\|^2$, portanto:

$$P^t P = \begin{bmatrix} \|v_1\|^2 & 0 & \dots & 0 \\ 0 & \|v_2\|^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \|v_n\|^2 \end{bmatrix}$$

Como P é uma matriz de autovetores ortogonais de A , visto que A é uma matriz simétrica, então define-se Q como uma matriz dos autovetores normalizados de A , então:

$$q_1 = \frac{v_1}{\|v_1\|}, \quad q_2 = \frac{v_2}{\|v_2\|}, \quad \dots \quad q_n = \frac{v_n}{\|v_n\|}$$

$$Q^t Q = \begin{bmatrix} \|q_1\|^2 & 0 & \dots & 0 \\ 0 & \|q_2\|^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \|q_n\|^2 \end{bmatrix}$$

Como $\|q_i\| = 1, \forall i$, tem-se:

$$Q^t Q = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

$$Q^t Q = I$$

Pós-multiplicando-se por Q^{-1} tem-se,

$$Q^t Q Q^{-1} = I Q^{-1}$$

$$Q^t = Q^{-1}$$

Assim, a matriz Q é ortogonal. Então, pode-se diagonalizar A utilizando-se Q ao invés de P , obtendo-se assim,

$$AQ = QD.$$

Portanto, analogamente aos resultados anteriores, obtém-se a diagonalização ou a decomposição da matriz A , respectivamente como:

$$Q^t A Q = D \text{ ou } A = Q D Q^t.$$

A expressão $A = Q D Q^t$ representa a decomposição da matriz A em autovalores e autovetores ortogonais e normalizados.

Portanto, uma matriz A é simétrica se, e somente se, pode ser decomposta em $A = QDQ^t$, onde Q é uma matriz ortogonal.

A denominação de **decomposição espectral** de A é devido a seguinte expressão,

$$A = QDQ^t = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^t \\ \mathbf{q}_2^t \\ \vdots \\ \mathbf{q}_n^t \end{bmatrix}$$

$$A = [\lambda_1 \mathbf{q}_1 \quad \lambda_2 \mathbf{q}_2 \quad \cdots \quad \lambda_n \mathbf{q}_n] \begin{bmatrix} \mathbf{q}_1^t \\ \mathbf{q}_2^t \\ \vdots \\ \mathbf{q}_n^t \end{bmatrix}$$

$$A = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^t + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^t + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^t,$$

em que cada um dos termos $\lambda_i \mathbf{q}_i \mathbf{q}_i^t$ é uma matriz de posto 1, sendo $\mathbf{q}_i \mathbf{q}_i^t$ a matriz de projeção sobre o subespaço gerado por \mathbf{q}_i . Isto é, é uma matriz capaz de gerar as projeções de um vetor \mathbf{w} qualquer do \mathbb{R}^n na direção de \mathbf{q}_i ,

$$(\text{Proj } \mathbf{w})_{\mathbf{q}_i} = c_i \mathbf{q}_i = (\mathbf{w}^t \mathbf{q}_i) \mathbf{q}_i = (\mathbf{q}_i^t \mathbf{w}) \mathbf{q}_i = \mathbf{q}_i \mathbf{q}_i^t \mathbf{w}$$

Por isso, muitas vezes a decomposição espectral também é conhecida como teorema espectral na forma de projeções.

A demonstração no \mathbb{R}^3 a seguir pode ser facilmente generalizada para vetores no \mathbb{R}^n .

Seja $\mathbf{q}_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix}$ um autovetor de uma matriz A e seja $\mathbf{w} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ um vetor

qualquer do \mathbb{R}^3 . Então

$$(\mathbf{q}_i \mathbf{q}_i^t) \mathbf{w} = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} [x_i \quad y_i \quad z_i] \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$(\mathbf{q}_i \mathbf{q}_i^t) \mathbf{w} = \begin{bmatrix} x_i x_i & x_i y_i & x_i z_i \\ x_i y_i & y_i y_i & y_i z_i \\ x_i z_i & y_i z_i & z_i z_i \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$(\mathbf{q}_i \mathbf{q}_i^t) \mathbf{w} = \begin{bmatrix} x x_i x_i + y x_i y_i + z x_i z_i \\ x x_i y_i + y y_i y_i + z y_i z_i \\ x x_i z_i + y y_i z_i + z z_i z_i \end{bmatrix}$$

$$(\mathbf{q}_i \mathbf{q}_i^t) \mathbf{w} = \begin{bmatrix} (xx_i)x_i + (yx_i)y_i + (zx_i)z_i \\ (xy_i)x_i + (yy_i)y_i + (zy_i)z_i \\ (xz_i)x_i + (yz_i)y_i + (zz_i)z_i \end{bmatrix}$$

$$(\mathbf{q}_i \mathbf{q}_i^t) \mathbf{w} = \begin{bmatrix} xx_i & yx_i & zx_i \\ xy_i & yy_i & zy_i \\ xz_i & yz_i & zz_i \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix}$$

$$(\mathbf{q}_i \mathbf{q}_i^t) \mathbf{w} = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix}$$

$$(\mathbf{q}_i \mathbf{q}_i^t) \mathbf{w} = \mathbf{q}_i \mathbf{w}^t \mathbf{q}_i$$

Como $\mathbf{w}^t \mathbf{q}_i$ é um escalar,

$$(\mathbf{q}_i \mathbf{q}_i^t) \mathbf{w} = (\mathbf{w}^t \mathbf{q}_i) \mathbf{q}_i = (\text{proj } \mathbf{w})_{\mathbf{q}_i}$$

Outra maneira de visualizar as propriedades matriz de projeção $\mathbf{q}_i \mathbf{q}_i^t$ é exemplificada a seguir:

Seja $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ um operador linear tal que $T(\mathbf{w}) = A\mathbf{w}$, onde A é uma matriz simétrica, e $\beta = \{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$ uma base ortonormal de autovetores de A . Assim, \mathbf{w} pode ser escrito como combinação linear dos vetores de β .

$$\mathbf{w} = c_1 \mathbf{q}_1 + c_2 \mathbf{q}_2 + \dots + c_n \mathbf{q}_n$$

$$T(\mathbf{w}) = c_1 T(\mathbf{q}_1) + c_2 T(\mathbf{q}_2) + \dots + c_n T(\mathbf{q}_n)$$

$$T(\mathbf{w}) = \lambda_1 \mathbf{q}_1 c_1 + \lambda_2 \mathbf{q}_2 c_2 + \dots + \lambda_n \mathbf{q}_n c_n,$$

Recorde que $c_i = \mathbf{w}^t \mathbf{q}_i = \mathbf{q}_i^t \mathbf{w}$, portanto

$$T(\mathbf{w}) = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^t \mathbf{w} + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^t \mathbf{w} + \dots + \lambda_n \mathbf{q}_n \mathbf{q}_n^t \mathbf{w}$$

$$T(\mathbf{w}) = (\lambda_1 \mathbf{q}_1 \mathbf{q}_1^t + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^t + \dots + \lambda_n \mathbf{q}_n \mathbf{q}_n^t) \mathbf{w}$$

$$T(\mathbf{w}) = A\mathbf{w}$$

Exemplo detalhado de uma projeção no \mathbb{R}^3 utilizando-se decomposição espectral

Para ilustrar de forma concreta uma decomposição espectral será utilizada uma matriz $A_{3 \times 3}$ simétrica retirada de Lay (1999, p. 419) em que será

abordado o passo a passo para realização de uma decomposição espectral, conforme descrito a seguir:

- Passo 1: Determinar autovalores e autovetores da matriz A

$$\text{Seja } A = \begin{bmatrix} 3 & 2 & 0 \\ 2 & 2 & 2 \\ 0 & 2 & 1 \end{bmatrix}$$

$$A - \lambda I = \begin{bmatrix} 3 - \lambda & 2 & 0 \\ 2 & 2 - \lambda & 2 \\ 0 & 2 & 1 - \lambda \end{bmatrix}$$

Para encontrar os autovalores da matriz A , deve-se encontrar as raízes para a equação característica $\det(A - \lambda I) = 0$. Sendo assim, tem-se:

$$(3 - \lambda)(2 - \lambda)(1 - \lambda) - 4(1 - \lambda) - 4(3 - \lambda) = 0$$

$$(3 - \lambda)(2 - \lambda)(1 - \lambda) - 8(2 - \lambda) = 0$$

$$(2 - \lambda)[(3 - \lambda)(1 - \lambda) - 8] = 0$$

$$(2 - \lambda)(\lambda^2 - 4\lambda - 5) = 0$$

As raízes da equação ou autovalores são $\lambda_1 = 5$, $\lambda_2 = 2$, $\lambda_3 = -1$

Para encontrar um dos autovetores associados a $\lambda_1 = 5$, deve-se resolver o seguinte sistema: $(A - 5I)v = 0$ para uma solução não trivial, ou seja, $v \neq 0$. Então tem-se:

$$(A - 5I) = \begin{bmatrix} 3 - 5 & 2 & 0 \\ 2 & 2 - 5 & 2 \\ 0 & 2 & 1 - 5 \end{bmatrix} = \begin{bmatrix} -2 & 2 & 0 \\ 2 & -3 & 2 \\ 0 & 2 & -4 \end{bmatrix}$$

Escalonando:

$$\begin{bmatrix} -2 & 2 & 0 & \vdots & 0 \\ 2 & -3 & 2 & \vdots & 0 \\ 0 & 2 & -4 & \vdots & 0 \end{bmatrix} \sim \begin{bmatrix} -2 & 2 & 0 & \vdots & 0 \\ 0 & -1 & 2 & \vdots & 0 \\ 0 & 2 & -4 & \vdots & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 0 & \vdots & 0 \\ 0 & -1 & 2 & \vdots & 0 \\ 0 & 0 & 0 & \vdots & 0 \end{bmatrix}$$

O sistema equivalente obtido é dado por

$$\begin{cases} x - y = 0 \\ y - 2z = 0 \end{cases} \quad \therefore \quad x = y = 2z$$

Então,

$$(x, y, z) = (2z, 2z, z) = z(2, 2, 1) \therefore \mathbf{v}_1 = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$$

Para encontrar um dos autovetores associados a $\lambda_2 = 2$, deve-se resolver o seguinte sistema: $(A - 2I)\mathbf{v} = 0$ para uma solução não trivial, ou seja, $\mathbf{v} \neq 0$. Então tem-se:

$$(A - 2I) = \begin{bmatrix} 3-2 & 2 & 0 \\ 2 & 2-2 & 2 \\ 0 & 2 & 1-2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 0 & 2 \\ 0 & 2 & -1 \end{bmatrix}$$

Escalonando:

$$\begin{bmatrix} 1 & 2 & 0 & : & 0 \\ 2 & 0 & 2 & : & 0 \\ 0 & 2 & -1 & : & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 0 & : & 0 \\ 0 & -4 & 2 & : & 0 \\ 0 & 2 & -1 & : & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 0 & : & 0 \\ 0 & 2 & -1 & : & 0 \\ 0 & 0 & 0 & : & 0 \end{bmatrix}$$

$$\begin{cases} x + 2y = 0 \\ 2y - z = 0 \end{cases} \therefore -x = 2y = z$$

Então,

$$(x, y, z) = (-2y, y, 2y) = y(-2, 1, 2) \therefore \mathbf{v}_2 = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}$$

Para encontrar um dos autovetores associados a $\lambda_3 = -1$, deve-se resolver o seguinte sistema: $(A + I) \cdot \mathbf{v} = 0$ para uma solução não trivial, ou seja, $\mathbf{v} \neq 0$. Então tem-se:

$$(A + I) = \begin{bmatrix} 3+1 & 2 & 0 \\ 2 & 2+1 & 2 \\ 0 & 2 & 1+1 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 0 \\ 2 & 3 & 2 \\ 0 & 2 & 2 \end{bmatrix}$$

Escalonando:

$$\begin{bmatrix} 4 & 2 & 0 & : & 0 \\ 2 & 3 & 2 & : & 0 \\ 0 & 2 & 2 & : & 0 \end{bmatrix} \sim \begin{bmatrix} 2 & 1 & 0 & : & 0 \\ 0 & 2 & 2 & : & 0 \\ 0 & 2 & 2 & : & 0 \end{bmatrix} \sim \begin{bmatrix} 2 & 1 & 0 & : & 0 \\ 0 & 1 & 1 & : & 0 \\ 0 & 0 & 0 & : & 0 \end{bmatrix}$$

$$\begin{cases} 2x + y = 0 \\ y + z = 0 \end{cases} \therefore 2x = -y = z$$

Então,

$$(x, y, z) = (x, -2x, 2x) = x(1, -2, 2) \therefore v_3 = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$$

- Passo 2: Determinado as matrizes D e P

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix} \text{ em que } \lambda_1 \geq \lambda_2 \geq \lambda_3$$

$$P = [v_1 \quad v_2 \quad v_3] = \begin{bmatrix} 2 & -2 & 1 \\ 2 & 1 & -2 \\ 1 & 2 & 2 \end{bmatrix}$$

Então a matriz A pode ser decomposta como se segue:

$$A = PDP^{-1}$$
$$A = \begin{bmatrix} 2 & -2 & 1 \\ 2 & 1 & -2 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 2/9 & 2/9 & 1/9 \\ -2/9 & 1/9 & 2/9 \\ 1/9 & -2/9 & 2/9 \end{bmatrix}$$

- Passo 3: Determinando a matriz Q

Normalizando os autovetores v_1, v_2, v_3 , encontra-se os vetores q_1, q_2, q_3 , respectivamente. Tais vetores formarão uma nova base ortonormal

$$P = [v_1 \quad v_2 \quad v_3] = \begin{bmatrix} 2 & -2 & 1 \\ 2 & 1 & -2 \\ 1 & 2 & 2 \end{bmatrix}$$

Encontrando as normas dos autovetores v_1, v_2 e v_3 , tem-se:

$$\|v_1\| = \sqrt{2^2 + 2^2 + 1^2} = 3$$

$$\|v_2\| = \sqrt{(-2)^2 + 1^2 + 2^2} = 3$$

$$\|v_3\| = \sqrt{1^2 + (-2)^2 + 2^2} = 3$$

Encontrando os versores de v_1, v_2 e v_3 , determinam-se os vetores q_1, q_2 e q_3 , respectivamente conforme mostra-se a seguir.

$$\mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \frac{1}{3} \cdot \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 2/3 \\ 1/3 \end{bmatrix}$$

$$\mathbf{q}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \frac{1}{3} \cdot \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2/3 \\ 1/3 \\ 2/3 \end{bmatrix}$$

$$\mathbf{q}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \frac{1}{3} \cdot \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}$$

Portanto a matriz Q é formada por

$$Q = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{q}_3] = \begin{bmatrix} 2/3 & -2/3 & 1/3 \\ 2/3 & 1/3 & -2/3 \\ 1/3 & 2/3 & 2/3 \end{bmatrix}$$

Finalmente, determina-se a decomposição espectral de A :

$$A = QDQ^t$$

$$A = \begin{bmatrix} 2/3 & -2/3 & 1/3 \\ 2/3 & 1/3 & -2/3 \\ 1/3 & 2/3 & 2/3 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 2/3 & 2/3 & 1/3 \\ -2/3 & 1/3 & 2/3 \\ 1/3 & -2/3 & 2/3 \end{bmatrix}$$

$$A = \lambda_1 \mathbf{q}_1 \mathbf{q}_1^t + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^t + \lambda_3 \mathbf{q}_3 \mathbf{q}_3^t$$

$$A = 5 \begin{bmatrix} 4/9 & 4/9 & 2/9 \\ 4/9 & 4/9 & 2/9 \\ 2/9 & 2/9 & 1/9 \end{bmatrix} + 2 \begin{bmatrix} 4/9 & -2/9 & -4/9 \\ -2/9 & 1/9 & 2/9 \\ -4/9 & 2/9 & 4/9 \end{bmatrix} - 1 \begin{bmatrix} 1/9 & -2/9 & 2/9 \\ -2/9 & 4/9 & -4/9 \\ 2/9 & -4/9 & 4/9 \end{bmatrix}$$

As matrizes $\mathbf{q}_i \mathbf{q}_i^t$ descritas possuem posto 1, pois todas suas linhas são múltiplos escalares de \mathbf{q}_i e são denominadas de matrizes de projeções, visto que através delas pode-se projetar um vetor na direção de \mathbf{q}_i . Para exemplificar pode utilizar das matrizes $\mathbf{q}_i \mathbf{q}_i^t$ para projetar um vetor \mathbf{w} em direção da nova base ortonormal $\beta = \{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$, conforme apresentado na Figura 12.

$$\text{Dado um vetor } \mathbf{w} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

$$(\text{proj } \mathbf{w})_{\mathbf{q}_1} = \begin{bmatrix} 4/9 & 4/9 & 2/9 \\ 4/9 & 4/9 & 2/9 \\ 2/9 & 2/9 & 1/9 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/9 \\ -2/9 \\ -1/9 \end{bmatrix}$$

$$(\text{proj } \mathbf{w})_{q_2} = \begin{bmatrix} 4/9 & -2/9 & -4/9 \\ -2/9 & 1/9 & 2/9 \\ -4/9 & 2/9 & 4/9 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -8/9 \\ 4/9 \\ 8/9 \end{bmatrix}$$

$$(\text{proj } \mathbf{w})_{q_3} = \begin{bmatrix} 1/9 & -2/9 & 2/9 \\ -2/9 & 4/9 & -4/9 \\ 2/9 & -4/9 & 4/9 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/9 \\ -2/9 \\ 2/9 \end{bmatrix}$$

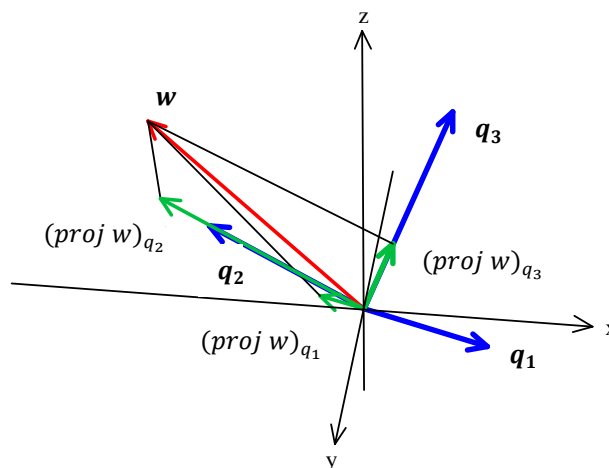


Figura 12 – Projeção do vetor \mathbf{w} sobre \mathbf{q}_1 , \mathbf{q}_2 e \mathbf{q}_3

2.1.4.3 Decomposição em valores singulares

Apresentou-se anteriormente algumas alternativas para se decompor uma matriz quadrada A . Porém, nem sempre as matrizes envolvidas em AC são quadradas, fazendo-se necessária uma nova forma de decomposição, que será apresentada nesta subseção.

Para decompor uma matriz $A_{m \times n}$ qualquer, existe um processo amplamente empregado na matemática e na estatística, denominado de decomposição em valores singulares (DVS). Uma adequada compreensão de DVS é um importante requisito para o entendimento da AC. Resumidamente, a DVS fornece $A = U\Sigma V^t$, sendo $U_{m \times m}$ e $V_{n \times n}$ matrizes ortogonais cujos vetores coluna formam uma base do \mathbb{R}^m e do \mathbb{R}^n , respectivamente, e $\Sigma_{m \times n}$ uma matriz (“semelhante a uma diagonal em blocos”), que contém a matriz diagonal D contendo as raízes quadradas dos autovalores ordenados de $A^t A$. A seguir detalha-se todo o processo da DVS.

O processo de DVS considera inicialmente a matriz $A^t A$ simétrica, e o importante resultado de que ela possui todos os autovalores reais e não negativos, conforme demonstrado a seguir.

Seja \mathbf{v} um autovetor unitário de $A^t A$, associado a algum λ , logo, $(A^t A)\mathbf{v} = \lambda\mathbf{v}$. Se λ é um autovalor de $A^t A$, então $\lambda \geq 0$, pois:

$$\begin{aligned}
 0 \leq \|A\mathbf{v}\|^2 &= \langle A\mathbf{v}, A\mathbf{v} \rangle \text{ (produto interno)} \\
 &= (A\mathbf{v})^t(A\mathbf{v}) \text{ (reescrevendo na forma matricial)} \\
 &= \mathbf{v}^t A^t A \mathbf{v} \\
 &= \mathbf{v}^t \lambda \mathbf{v} \\
 &= \lambda \mathbf{v}^t \mathbf{v} \\
 &= \lambda \|\mathbf{v}\|^2 \\
 &= \lambda
 \end{aligned}$$

A expressão valores singulares surge quando se considera as raízes quadradas dos autovalores de $A^t A$. Isto é, $0 \leq \|A\mathbf{v}\| = \sqrt{\lambda} = \sigma$, definem os valores singulares $\sigma_1, \sigma_2, \dots, \sigma_n$, que por convenção são ordenados de modo que $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$.

Para um melhor entendimento dos conceitos ora apresentados será realizada a decomposição de uma matriz associada à transformação linear $T: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ (POOLE, 2004).

Seja A uma matriz de transformação $T: \mathbb{R}^2 \rightarrow \mathbb{R}^3$, tal que $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Considere um vetor \mathbf{v} tal que $T(\mathbf{v}) = A\mathbf{v} = \mathbf{w}$,

$$T(x, y) = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + y \\ x \\ y \end{bmatrix}$$

Portanto, na transformação descrita acima x e y podem assumir qualquer valor. Porém, se for estabelecido à condição que o vetor \mathbf{v} tenha norma unitária, ou seja, $\|\mathbf{v}\| = 1$, percebe-se que as coordenadas x e y estão sujeitas à seguinte restrição:

$$\begin{aligned}
 \|\mathbf{v}\| &= \sqrt{x^2 + y^2} = 1 \\
 x^2 + y^2 &= 1.
 \end{aligned}$$

Sendo assim, as coordenadas x e y do vetor \mathbf{v} podem assumir qualquer valor desde que a equação $x^2 + y^2 = 1$ seja respeitada, o que gera

infinitos vetores \mathbf{v} sobre o círculo com raio igual a 1 no plano, ou seja, no \mathbb{R}^2 . Graficamente tem-se,

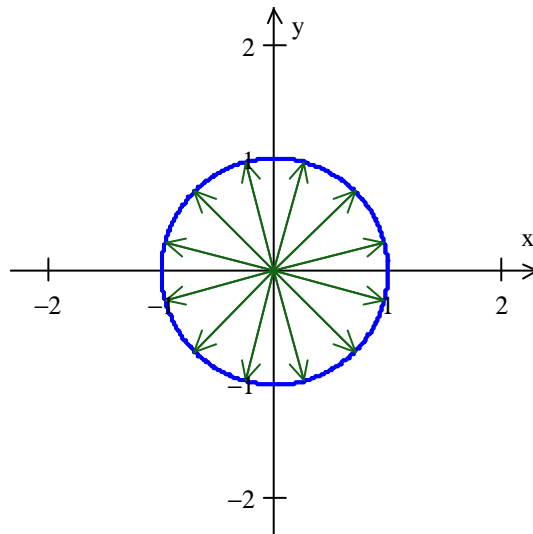


Figura 13 – Visualização de algumas possibilidades do vetor \mathbf{v} com a restrição $\|\mathbf{v}\| = 1$

Cada um dos infinitos vetores \mathbf{v} pertencentes ao \mathbb{R}^2 possui um vetor a ele associado no \mathbb{R}^3 , que pode ser determinado pela transformação linear $T(\mathbf{v}) = A\mathbf{v} = \mathbf{w}$. A norma do vetor $A\mathbf{v}$ pode ser então determinada como se segue:

$$\|A\mathbf{v}\|^2 = (A\mathbf{v})^t(A\mathbf{v}) = \mathbf{v}^t A^t A \mathbf{v} = \mathbf{v}^t \lambda \mathbf{v} = \lambda$$

Portanto,

$$\|A\mathbf{v}\| = \sqrt{\lambda} = \sigma$$

Pode-se perceber no desenvolvimento anterior que só foi possível determinar a norma do vetor $A\mathbf{v}$ após calcular os autovalores da matriz $A^t A$, que é simétrica de ordem 2x2 e portanto pode ser diagonalizada pelo teorema espectral $A^t A = Q D_\lambda Q^t$ (observe que na decomposição espectral a matriz D_λ não utiliza a raiz quadrada dos autovalores).

$$A^t A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$A^t A = Q D Q^t$$

$$A^t A = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Assim $\lambda_1 = 3$ e $\lambda_2 = 1$ e os autovetores unitários associados a estes autovalores são respectivamente

$$\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \quad \text{e} \quad \mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix},$$

que por sua vez formam uma matriz Q ortogonal. Fazendo $V = Q$ tem-se que

$$V = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \text{ cuja transposta é dada por } V^t = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}.$$

Então \mathbf{v}_1 e \mathbf{v}_2 formam uma base ortonormal do \mathbb{R}^2 que, por sua vez, podem ser transformados em $A\mathbf{v}_1$ e $A\mathbf{v}_2$. Estes, depois de normalizados, integrarão uma base ortonormal do \mathbb{R}^3 . Pelo processo de ortogonalização de Gram-Schmidt é possível determinar um terceiro vetor que seja unitário e ortogonal a $A\mathbf{v}_1$ e $A\mathbf{v}_2$ para formar uma base do \mathbb{R}^3 .

$$A\mathbf{v}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 2/\sqrt{2} \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$A\mathbf{v}_2 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

Como já demonstrado que $\|A\mathbf{v}\| = \sqrt{\lambda} = \sigma$, da normalização de $A\mathbf{v}_1$ e $A\mathbf{v}_2$ obtém-se os vetores \mathbf{u}_1 e \mathbf{u}_2 respectivamente

$$\mathbf{u}_i = \frac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|} = \frac{A\mathbf{v}_i}{\sqrt{\lambda_i}} = \frac{1}{\sigma} A\mathbf{v}_i$$

$$\mathbf{u}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 2/\sqrt{2} \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{1} \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

Finalmente, \mathbf{u}_3 é obtido pelo procedimento de ortogonalização de Gram-Schmidt,

$$\mathbf{u}_3 = \begin{bmatrix} -1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$$

Com os vetores \mathbf{u}_1 , \mathbf{u}_2 e \mathbf{u}_3 é possível determinar-se a matriz ortogonal U ;

$$U = \begin{bmatrix} 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix}$$

A matriz Σ que contém os valores singulares A é formada pela raiz quadrada da matriz diagonal D_λ tendo sido excluídos os autovalores iguais à zero, e completada com matrizes nulas de tamanho apropriado para que Σ seja de ordem $m \times n$

$$\Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Assim A pode ser decomposta pela DVS como $A = U\Sigma V^t$.

$$A = U\Sigma V^t = \begin{bmatrix} 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}.$$

Na Figura 14, pode-se visualizar os vetores v_1 e v_2 que formam uma base em torno de um círculo no \mathbb{R}^2 e os vetores u_1 , u_2 , e u_3 que são ortogonais e formam uma base no \mathbb{R}^3 .

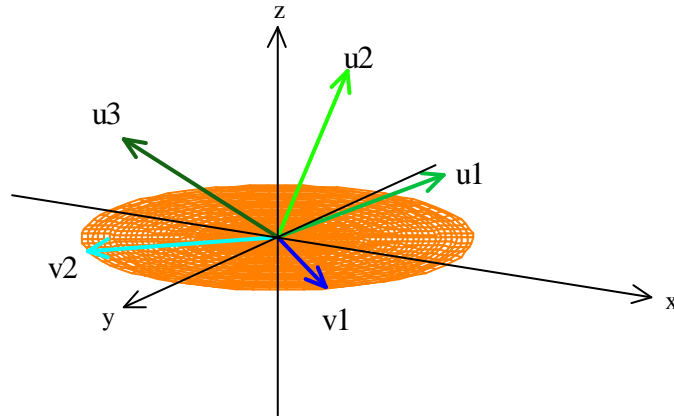


Figura 14 – Vetores que formam uma base em \mathbb{R}^2 e no \mathbb{R}^3

Resumindo os passos da decomposição em valores singulares da matriz A de ordem $m \times n$ e posto k , em $A = U\Sigma V^t$. Primeiramente determina-se a matriz $A^t A$, que por ser simétrica permite fazer a decomposição espectral em $A^t A = QD_\lambda Q^t$; faz-se $V = Q$ e determinam-se os vetores de U a partir de $\frac{Av_i}{\|Av_i\|} = u_i$, e completando-os, se necessário, até obter uma base ortonormal compatível com sua dimensão; determina-se a matriz diagonal D utilizando as k primeiras raízes quadradas dos autovalores $A^t A$ colocadas em ordem decrescente; obtém-se a matriz em blocos Σ composta pela matriz D e completada por matrizes nulas para que tenha ordem $m \times n$; finalmente obtém-se a decomposição em valores singulares da matriz A , por $A = U\Sigma V^t$.

É importante verificar as ordens das matrizes envolvidas na DVS: A é uma matriz $m \times n$; $A^t A$ resulta em uma matriz quadrada $n \times n$, e, portanto, Q e D_λ também o são; V também é uma matriz $n \times n$, U é uma matriz $m \times m$ e Σ é uma matriz $m \times n$. Como $\text{posto}(A) = k \leq \min\{m, n\}$, então D terá ordem $k \times k$.

Particiona-se U e V em submatrizes de modo que o primeiro bloco de cada uma contenha k vetores relativos aos k autovalores diferentes de zero, esquematicamente pode-se representar a DVS de A por:

$$A = [U_k \quad U_{m-k}] \begin{bmatrix} D_k & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_k^t \\ V_{n-k}^t \end{bmatrix}$$

sendo U_k uma matriz com k vetores de m coordenadas e V_k com k vetores de n coordenadas. Tem-se, então, que

$$A = [U_k D_k + U_{m-k} 0 \quad U_k 0 + U_{m-k} 0] \begin{bmatrix} V_k^t \\ V_{n-k}^t \end{bmatrix}$$

$$A = [U_k D_k \quad 0] \begin{bmatrix} V_k^t \\ V_{n-k}^t \end{bmatrix}$$

$$A = [U_k D_k V_k^t + V_{n-k}^t 0]$$

$$A = U_k D_k V_k^t$$

Esta última forma de escrever a matriz A é a DVS reduzida.

Usando a DVS reduzida pode-se obter um tipo de inversa generalizada para a matriz A quando esta não é quadrada ou quando é quadrada singular:

$$A^+ = V_k D_k^{-1} U_k^t$$

conhecida como inversa de Moore-Penrose, em que $AA^+A = A$; $A^+AA^+ = A^+$; $(AA^+)^t = AA^+$ e $(A^+A)^t = A^+A$ (REGAZZI, 2010).

2.1.5 Ajuste de quadrados mínimos

O método dos mínimos quadrados é uma técnica matemática amplamente utilizada na estatística para fazer o ajuste de um modelo de regressão a um conjunto de dados. É comum uma representação geométrica deste ajuste através de um gráfico no \mathbb{R}^2 como o mostrado na Figura 15, em que se procura minimizar a soma dos quadrados dos erros. Porém, utilizando-se o enfoque vetorial pode-se ter outra visão do problema, envolvendo a relação entre erro quadrático, projeção ortogonal e formas quadráticas que será de fundamental importância na determinação de um subespaço ótimo, necessário para redução de dimensionalidade da AC.

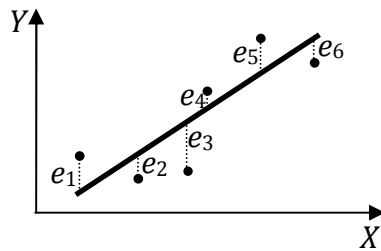


Figura 15 – Visão geométrica de um ajuste de quadrados mínimos

2.1.5.1 Projeção ortogonal e erro quadrático

Considere a situação apresentada pela

Figura 16, com projeções do vetor \mathbf{u} sobre o plano α em várias direções, sendo $\hat{\mathbf{u}}$ a projeção ortogonal de \mathbf{u} sobre o plano α ($\hat{\mathbf{u}} = (\text{Proj } \mathbf{u})_\alpha$), isto é, projeção na qual o plano projetor de \mathbf{u} é perpendicular ao plano α .

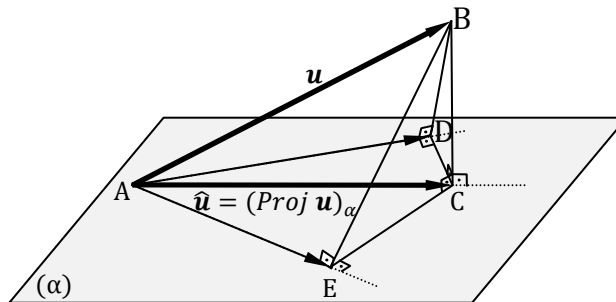


Figura 16 – Projeções do vetor \mathbf{u} sobre o plano α , sendo $\hat{\mathbf{u}}$ a projeção ortogonal

Esta figura mostra claramente algumas propriedades das projeções ortogonais:

- (i) A norma de \mathbf{u} , $\|\mathbf{u}\|$, é sempre maior que a norma de qualquer outra projeção, visto que os triângulos ABC, ABD e ABE são todos retângulos com hipotenusa \overline{AB} .
- (ii) A norma da projeção ortogonal, $\|\hat{\mathbf{u}}\|$, é maior que a norma de qualquer outra projeção, visto que os triângulos ACD e ACE são retângulos com hipotenusa \overline{AC} .
- (iii) A distância do ponto B ao plano α é dada pela distância do ponto B ao ponto C, isto é, $\text{dist}(B, \alpha) = \text{dist}(B, C)$.
- (iv) Considerando $\text{dist}(B, C)$, $\text{dist}(B, D)$ e $\text{dist}(B, E)$, a menor delas é $\text{dist}(B, C)$, pois nos triângulos retângulos BCD e BCE as hipotenusas são \overline{BD} e \overline{BE} respectivamente, sendo \overline{BC} um cateto.

(v) Tem-se que $dist(B,C) = \|\overrightarrow{CB}\|$ e $\overrightarrow{CB} = \overrightarrow{AB} - \overrightarrow{AC} = \mathbf{u} - \hat{\mathbf{u}}$, que é ortogonal ao plano α .

Logo, pode-se relacionar que a maior projeção de um vetor sobre o plano α gera a menor diferença entre o vetor e sua projeção, isto é, a projeção ortogonal minimiza a diferença entre um vetor e sua projeção, que, por consequência, maximiza a projeção de um vetor sobre um plano.

A diferença $\mathbf{u} - \hat{\mathbf{u}}$ pode ser vista como um vetor de erro, isto é, como o erro cometido ao se utilizar a projeção de \mathbf{u} no lugar de \mathbf{u} , e minimizar o erro significa minimizar a norma dessa diferença:

$$\|\mathbf{e}\| = \|\mathbf{u} - \hat{\mathbf{u}}\| \Rightarrow \|\mathbf{e}\|^2 = \|\mathbf{u} - \hat{\mathbf{u}}\|^2.$$

Se \mathbf{u} e $\hat{\mathbf{u}}$ têm coordenadas $\mathbf{u} = [u_1 \ u_2 \ \dots \ u_n]^t$ e $\hat{\mathbf{u}} = [\hat{u}_1 \ \hat{u}_2 \ \dots \ \hat{u}_n]^t$, então,

$$\|\mathbf{e}\|^2 = \|\mathbf{u} - \hat{\mathbf{u}}\|^2 = (u_1 - \hat{u}_1)^2 + (u_2 - \hat{u}_2)^2 + \dots + (u_n - \hat{u}_n)^2.$$

Assim, minimizar o erro é o mesmo que minimizar o erro quadrático que é o mesmo que minimizar a soma de quadrados das coordenadas do erro.

A seguir será usado um exemplo de Matos e Amaral (2008, aula 33, p.6)

Seja W um subespaço do \mathbb{R}^3 com base ortogonal $\{[2 \ 2 \ 0]^t, [-2 \ 2 \ 0]^t\}$ e o vetor $\mathbf{u} = [0 \ 2 \ 2]^t$.

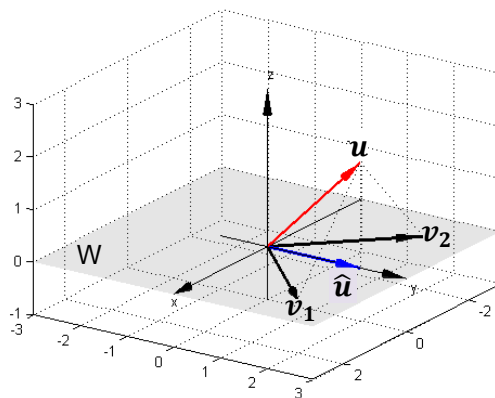


Figura 17 – Projeção do vetor \mathbf{u} no subespaço W

Fonte: Matos e Amaral, 2008, p. 2.

Vê-se pela Figura 17 que o vetor \mathbf{u} não pertence ao subespaço W , então, a melhor representação de \mathbf{u} em W é dada por sua projeção ortogonal. Sendo W gerado por $\mathbf{v}_1 = [2 \ 2 \ 0]^t$ e por $\mathbf{v}_2 = [-2 \ 2 \ 0]^t$ que são ortogonais,

mas não são vetores normais, é preciso normalizá-los primeiro para depois determinar as projeções.

$$\mathbf{v}'_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \frac{[2 \ 2 \ 0]^t}{\sqrt{8}}$$

$$\mathbf{v}'_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \frac{[-2 \ 2 \ 0]^t}{\sqrt{8}}$$

$$\hat{\mathbf{u}} = (\text{Proj } \mathbf{u})_W = (\text{Proj } \mathbf{u})_{v_1} + (\text{Proj } \mathbf{u})_{v_2}$$

$$\hat{\mathbf{u}} = (\mathbf{u}^t \mathbf{v}'_1) \mathbf{v}'_1 + (\mathbf{u}^t \mathbf{v}'_2) \mathbf{v}'_2$$

$$\hat{\mathbf{u}} = [0 \ 2 \ 2] \cdot \frac{1}{\sqrt{8}} \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix} \cdot \frac{1}{\sqrt{8}} \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix} + [0 \ 2 \ 2] \cdot \frac{1}{\sqrt{8}} \begin{bmatrix} -2 \\ 2 \\ 0 \end{bmatrix} \cdot \frac{1}{\sqrt{8}} \begin{bmatrix} -2 \\ 2 \\ 0 \end{bmatrix}$$

$$\hat{\mathbf{u}} = \frac{1}{8} \cdot 4 \cdot \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix} + \frac{1}{8} \cdot 4 \cdot \begin{bmatrix} -2 \\ 2 \\ 0 \end{bmatrix}$$

$$\hat{\mathbf{u}} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

$$\hat{\mathbf{u}} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$$

O vetor $\hat{\mathbf{u}}$ pertencente a W é a melhor aproximação do vetor \mathbf{u} neste subespaço, com vetor de erro e sua norma quadrática dada por:

$$\mathbf{e} = \mathbf{u} - \hat{\mathbf{u}}$$

$$\mathbf{e} = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix} - \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} \Rightarrow \|\mathbf{e}\| = 2 \Rightarrow \|\mathbf{e}\|^2 = 4$$

O subespaço W também pode ser gerado por outra base de vetores ortogonais $\mathbf{v}_3 = [1 \ 3 \ 0]^t$ e $\mathbf{v}_4 = [3 \ -1 \ 0]^t$. Determinando a projeção de \mathbf{u} em W através dos vetores \mathbf{v}_3 e \mathbf{v}_4 tem-se:

$$\mathbf{v}'_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \frac{[1 \ 3 \ 0]^t}{\sqrt{10}}$$

$$\mathbf{v}'_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \frac{[3 \ -1 \ 0]^t}{\sqrt{10}}$$

$$\hat{\mathbf{u}} = (\text{Proj } \mathbf{u})_W = (\text{Proj } \mathbf{u})_{v_3} + (\text{Proj } \mathbf{u})_{v_4}$$

$$\hat{\mathbf{u}} = (\mathbf{u}^t \mathbf{v}'_3) \mathbf{v}'_3 + (\mathbf{u}^t \mathbf{v}'_4) \mathbf{v}'_4$$

$$\hat{\mathbf{u}} = [0 \ 2 \ 2] \cdot \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} \cdot \frac{1}{\sqrt{10}} \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} + [0 \ 2 \ 2] \cdot \frac{1}{\sqrt{10}} \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix} \cdot \frac{1}{\sqrt{10}} \begin{bmatrix} -3 \\ -1 \\ 0 \end{bmatrix}$$

$$\hat{\mathbf{u}} = \frac{1}{10} \cdot 6 \cdot \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} + \frac{1}{10} \cdot (-2) \cdot \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix}$$

$$\hat{\mathbf{u}} = \begin{bmatrix} 3/5 \\ 9/5 \\ 0 \end{bmatrix} + \begin{bmatrix} -3/5 \\ 1/5 \\ 0 \end{bmatrix}$$

$$\hat{\mathbf{u}} = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$$

que é a mesma projeção determinada anteriormente pela outra base ortogonal de W . Este resultado mostra que a projeção de um vetor num subespaço não depende da base escolhida (MATOS e AMARAL, 2008, p.2).

2.1.6 Formas quadráticas

Forma quadrática é uma expressão em que cada um de seus termos possui grau total igual a 2, desta maneira:

- $5x^2 - 3y^2 + 2xy$ é uma forma quadrática,
- $3x^2 + 4y^2 - z^2 - 3xy$ é uma forma quadrática e
- $3x^2 + 4y^2 - 5x$ não é forma quadrática.

As expressões quadráticas podem ser representadas na forma matricial como mostrado a seguir:

$$ax^2 + by^2 + cxy = [x \ y] \begin{bmatrix} a & c/2 \\ c/2 & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{x}^t \mathbf{A} \mathbf{x}$$

$$\underbrace{ax^2 + by^2 + cz^2}_{\text{Termos não mistos}} + \underbrace{dxy + exz + fyz}_{\text{Termos mistos}} = [x \quad y \quad z] \begin{bmatrix} a & d/2 & e/2 \\ d/2 & b & f/2 \\ e/2 & f/2 & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{x}^t \mathbf{A} \mathbf{x}$$

Assim,

$$3x^2 - 5y^2 + 7z^2 - 4xy + xz = [x \quad y \quad z] \begin{bmatrix} 3 & -2 & 1/2 \\ -2 & -5 & 0 \\ 1/2 & 0 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{x}^t \mathbf{A} \mathbf{x}$$

$$3x^2 - 5y^2 + 7z^2 = [x \quad y \quad z] \begin{bmatrix} 3 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{x}^t \mathbf{A} \mathbf{x}$$

Verifica-se pelas expressões anteriores que a matriz A é sempre simétrica, sendo que quando a expressão não possui termos mistos a matriz A será uma matriz diagonal, pois os coeficientes dos termos mistos serão iguais à zero.

A definição formal para forma quadrática aparece em Poole (2004, p. 368).

Definição: Uma forma quadrática em n variáveis é uma função: $f: \mathbb{R}^n \rightarrow \mathbb{R}$ da forma $f(x) = \mathbf{x}^t \mathbf{A} \mathbf{x}$ onde A é uma matriz simétrica $n \times n$ e x está em \mathbb{R}^n . A matriz A é a matriz associada a f .

A Figura 18 apresenta gráficos de várias formas quadráticas e que por não possuírem termos mistos estão centradas na origem, Segundo Poole (2004) as formas quadráticas que não possuem termos mistos são mais facilmente analisadas.

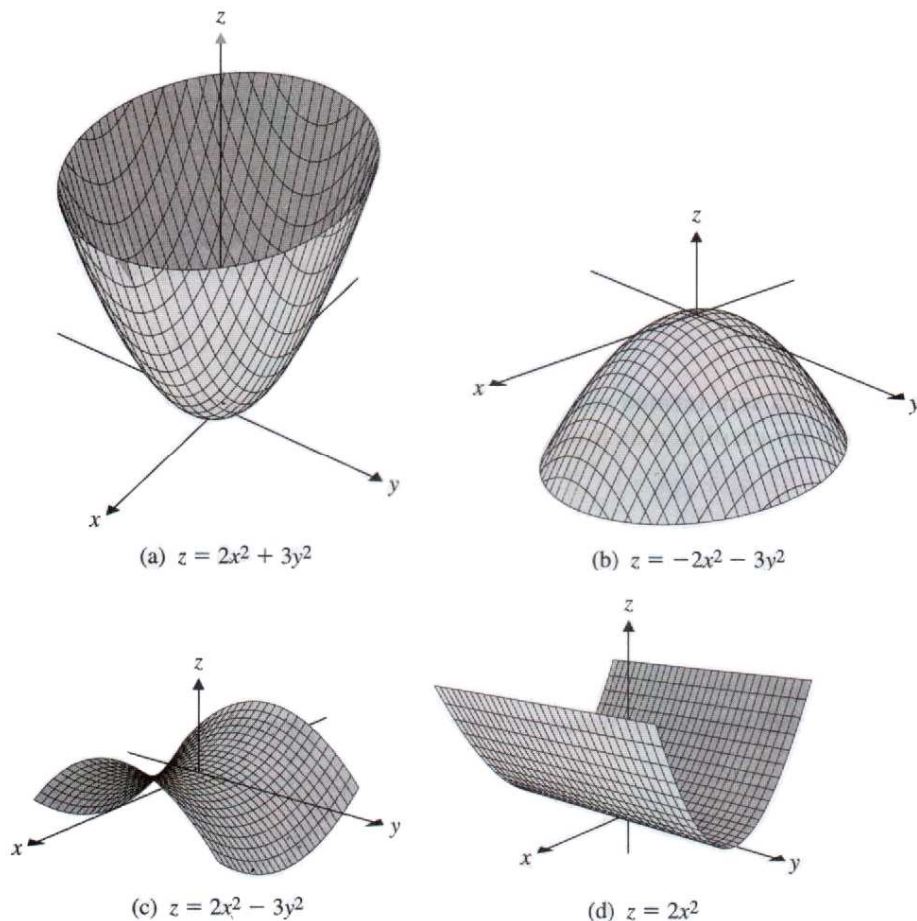


Figura 18 – Gráficos das formas quadráticas $f(x, y)$

Fonte: Poole, 2004, p. 369.

Então, a análise de formas quadráticas com termos mistos pode ser mais simples se for possível obter uma transformação no vetor x de tal forma que a matriz A possa ser diagonalizada.

Como a matriz A na forma quadrática é sempre simétrica, então pode-se utilizar o teorema espectral para diagonalizar a matriz A , que associado a uma transformação do vetor x permite eliminar os termos mistos da função, preservando a mesma forma quadrática. A seguir descreve-se o processo para eliminação dos termos mistos.

- Forma quadrática com termos mistos:

$$f(x) = x^t A x$$

- Diagonalização da matriz A

$$A = Q D Q^t \Rightarrow D = Q^T A Q$$

- Transformação do vetor x

$$x = Qy \Rightarrow y = Q^{-1}x = Q^t x$$

- Forma quadrática sem termos mistos

$$\begin{aligned} f(x) &= x^t A x \\ &= (Qy)^t A (Qy) \\ &= y^t Q^t (QDQ^t) Qy \\ &= y^t D y \\ &= y^t D y = g(y) \end{aligned}$$

Concluindo, $f(x) = g(y)$, fazendo $y = Q^t x$

$$x^t A x = y^t D y = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_n y_n^2$$

Esta é uma forma quadrática sem termos mistos na qual os coeficientes de cada termo são autovalores da matriz A . Este processo é denominado de diagonalização de formas quadráticas, também conhecido como teorema dos eixos principais.

Algumas consequências interessantes do teorema dos eixos principais:

- a) Se $y = Q^t x$ e $\|x\| = 1$ então $\|y\| = 1$, ou seja, a transformação de um vetor unitário é também um vetor unitário.

$$\begin{aligned} \|y\|^2 &= y^t y = (Q^t x)^t (Q^t x) \\ &= x^t Q Q^t x \\ &= x^t x \\ &= \|x\|^2 = 1 \end{aligned}$$

- b) $\lambda_1 \geq f(x) \geq \lambda_n$

Lembrando que se

$$\lambda_2 \leq \lambda_1,$$

então

$$\lambda_2 y_1^2 \leq \lambda_1 y_1^2$$

Assim,

$$\begin{aligned} \mathbf{x}^t \mathbf{A} \mathbf{x} &= \mathbf{y}^t \mathbf{D} \mathbf{y} \\ &= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \\ &\leq \lambda_1 y_1^2 + \lambda_1 y_2^2 + \cdots + \lambda_1 y_n^2 \\ &\leq \lambda_1 (y_1^2 + y_2^2 + \cdots + y_n^2) \\ &\leq \lambda_1 \|\mathbf{y}\|^2 \\ &\leq \lambda_1 \end{aligned}$$

A mesma ideia pode ser aplicada para

$$\begin{aligned} \lambda_n &\leq \lambda_{n-1} \\ \lambda_n y_n^2 &\leq \lambda_{n-1} y_n^2 \end{aligned}$$

Portanto,

$$\begin{aligned} \mathbf{x}^t \mathbf{A} \mathbf{x} &= \mathbf{y}^t \mathbf{D} \mathbf{y} \\ &= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \\ &\geq \lambda_n y_1^2 + \lambda_n y_2^2 + \cdots + \lambda_n y_n^2 \\ &\geq \lambda_n (y_1^2 + y_2^2 + \cdots + y_n^2) \\ &\geq \lambda_n \|\mathbf{y}\|^2 \\ &\geq \lambda_n \end{aligned}$$

- c) O valor máximo de $f(x)$ é λ_1 , quando x é um autovetor unitário associado a λ_1

Se V_{λ_1} é um subespaço formado pelos infinitos autovetores associados a λ_1 e um vetor $\mathbf{q}_1 \in V_{\lambda_1}$ com $\|\mathbf{q}_1\| = \|\mathbf{q}_1\|^2 = 1$, então

$$\begin{aligned} f(\mathbf{q}_1) &= \mathbf{q}_1^t \mathbf{A} \mathbf{q}_1 \\ f(\mathbf{q}_1) &= \mathbf{q}_1^t \lambda_1 \mathbf{q}_1 \\ f(\mathbf{q}_1) &= \lambda_1 \mathbf{q}_1^t \mathbf{q}_1 \\ f(\mathbf{q}_1) &= \lambda_1 \|\mathbf{q}_1\|^2 \\ f(\mathbf{q}_1) &= \lambda_1 \end{aligned}$$

- d) O valor mínimo de $f(x)$ é λ_n , quando x é um autovetor unitário associado a λ_n

Se V_{λ_n} é um subespaço formado pelos infinitos autovetores associados a λ_n e um vetor $\mathbf{q}_n \in V_{\lambda_n}$ com $\|\mathbf{q}_n\| = \|\mathbf{q}_n\|^2 = 1$, então

$$\begin{aligned}
f(\mathbf{q}_n) &= \mathbf{q}_n^t A \mathbf{q}_n \\
f(\mathbf{q}_n) &= \mathbf{q}_n^t \lambda_n \mathbf{q}_n \\
f(\mathbf{q}_n) &= \lambda_n \mathbf{q}_n^t \mathbf{q}_n \\
f(\mathbf{q}_n) &= \lambda_n \|\mathbf{q}_n\|^2 \\
f(\mathbf{q}_n) &= \lambda_n
\end{aligned}$$

2.1.7 Subespaço ótimo

A determinação de um subespaço ótimo constitui a ideia central da AC, que é obter um subespaço de dimensão reduzida que maximize as projeções de um conjunto de vetores.

Seja A uma matriz de ordem $m \times n$ formada por um conjunto de m vetores-linha de um espaço vetorial de dimensão n , isto é, de vetores pertencentes ao \mathbb{R}^n , sendo $\text{posto}(A) = k$. Deseja-se obter um vetor unitário x do mesmo espaço vetorial de tal maneira que maximize as projeções dos vetores-linha de A sobre si.

Maximizar as projeções é o mesmo que minimizar o erro cometido ao tomar as projeções dos vetores-linha de A sobre x em substituição aos próprios vetores-linha. Como já visto, a projeção que minimiza os erros é a ortogonal, cuja norma pode ser obtida diretamente pelo produto interno entre um vetor-linha e o vetor unitário x . Fazendo-se, então, Ax obtém-se um vetor com as normas das projeções ortogonais dos vetores-linha de A sobre x :

$$Ax = \begin{bmatrix} \boxed{l_1} \\ \boxed{l_2} \\ \vdots \\ \boxed{l_m} \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} l_1 \cdot x \\ l_2 \cdot x \\ \vdots \\ l_m \cdot x \end{bmatrix} = \begin{bmatrix} \|(Proj l_1)_x\| \\ \|(Proj l_2)_x\| \\ \vdots \\ \|(Proj l_m)_x\| \end{bmatrix}$$

Ao se fazer $(Ax)^t Ax$ obtém-se o somatório das normas quadráticas:

$$(Ax)^t Ax = [l_1 \cdot x \quad l_2 \cdot x \quad \cdots \quad l_m \cdot x] \begin{bmatrix} l_1 \cdot x \\ l_2 \cdot x \\ \vdots \\ l_m \cdot x \end{bmatrix}$$

$$(A\mathbf{x})^t A\mathbf{x} = \|(Proj \mathbf{l}_1)_x\|^2 + \|(Proj \mathbf{l}_2)_x\|^2 + \dots + \|(Proj \mathbf{l}_m)_x\|^2.$$

Mas,

$$(A\mathbf{x})^t A\mathbf{x} = \mathbf{x}^t (A^t A)\mathbf{x} = \sum_{i=1}^m \|(Proj \mathbf{l}_i)_x\|^2$$

que por ser uma forma quadrática – $A^t A$ é uma matriz simétrica – pode ser maximizada fazendo uso do teorema dos eixos principais, como visto no item anterior.

Para maximizar a forma quadrática faz-se necessária sua diagonalização, por meio da decomposição espectral, obtendo-se a matriz Q de autovetores e D_λ de autovalores. O maior autovalor, designado por λ_1 , representa o maior valor que a função quadrática pode ter, isto é, no caso presente este valor representa o maior somatório das projeções quadráticas. O autovetor associado a λ_1 , q_1 , é o vetor unitário que maximiza as projeções dos vetores-linha de A .

Para formar uma base ortonormal do subespaço linha de A serão necessários k vetores, pois $posto(A) = k$. Assim, desejando um segundo vetor unitário, ortogonal ao primeiro e que maximize as projeções das linhas de A sobre si, terá de ser tomado o vetor q_2 correspondente a λ_2 , que é o segundo maior valor que o somatório das normas quadráticas pode assumir. Desejando um terceiro vetor unitário, ortogonal a q_1 e q_2 e que maximize as projeções das linhas de A sobre si, terá de ser tomado o vetor q_3 correspondente a λ_3 , que é o terceiro maior valor que o somatório das normas quadráticas pode assumir. E assim por diante até o k -ésimo vetor q_k correspondente a λ_k . Visto que $posto(A) = posto(D_\lambda)$, os demais autovalores são obrigatoriamente iguais à zero, levando a projeções nulas que, portanto, não interessam.

Desejando-se projetar os vetores-linha de A num subespaço de dimensão r menor que a dimensão do espaço linha de A , isto é, $r < k$, basta tomar os r autovetores associados aos r primeiros maiores autovalores de $A^t A$. Estes autovetores formarão a base ortonormal do subespaço ótimo de dimensão r .

Outro enfoque para a definição dos eixos principais, a partir de multiplicadores de Lagrange, pode ser visto em Pamplona (1998) e Ferreira (2008).

Será, agora, feito um exemplo, considerando uma matriz A de dimensão 5×6 , e $\text{posto}(A) = 4$

$$A = \begin{bmatrix} 1 & 2 & -1 & 0 & 0 & 0 \\ -2 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 2 & 2 & 1 & 1 \\ 1 & 2 & 0 & 0 & 0 & 1 \\ -2 & 3 & 1 & 0 & 1 & 2 \end{bmatrix}$$

Sendo $A^t A = Q D_\lambda Q^t$, com

$$Q = \begin{bmatrix} 0,29 & 0,36 & 0,80 & 0,27 & 0 & -0,25 \\ -0,81 & 0,43 & 0,16 & -0,16 & -0,31 & 0,02 \\ -0,14 & -0,61 & 0,14 & 0,38 & -0,63 & -0,20 \\ 0,02 & -0,40 & 0,52 & -0,56 & 0 & 0,50 \\ -0,19 & -0,27 & 0,09 & -0,42 & 0,32 & -0,78 \\ -0,45 & -0,26 & 0,17 & 0,51 & 0,63 & 0,20 \end{bmatrix}$$

$$D_\lambda = \begin{bmatrix} 22,95 & 0 & 0 & 0 & 0 & 0 \\ 0 & 14,12 & 0 & 0 & 0 & 0 \\ 0 & 0 & 9,51 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,41 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Desejando um subespaço ótimo de dimensão 2, basta tomar $W = [(0,29; -0,81; -0,14; 0,02; -0,19; -0,45), (0,36; 0,43; -0,61; -0,40; -0,27; -0,26)]$

Como o somatório geral das projeções quadráticas nos 4 eixos principais é dado por $22,95 + 14,12 + 9,51 + 0,41 = 46,99$, então o subespaço ótimo W de dimensão 2 captura $\frac{22,95 + 14,12}{46,99} \times 100 = 78,9\%$ do somatório geral.

2.2 Análise de correspondência: metodologia e um exemplo introdutório

O embasamento algébrico fornecido até o momento permitirá compreender os princípios matemáticos e estatísticos da AC, tais como definição de massa, perfis, nuvens de pontos, distância de qui-quadrado, inércia, etc.

Esta seção abordará todas as fases para realização de uma AC simples, descrevendo-se inicialmente os fundamentos teóricos e em seguida detalhando-se todos os cálculos com um exemplo ilustrativo. Em cada uma dessas fases os cálculos são justificados e são introduzidos conceitos geométricos pertinentes, pois, como diz Naito (2007, p. 7), “Apesar dos conceitos geométricos constituírem os alicerces da maioria das técnicas

multivariadas, muitos livros de análise de dados multivariados apresentam apenas uma descrição sucinta destes conceitos”.

2.2.1 Tabela e matriz de dados

A análise de correspondência simples se inicia com a elaboração de uma tabela de contingência de dupla entrada com as frequências absolutas de duas variáveis categóricas, conforme mostrado na Tabela 1:

Tabela 1 – Estrutura geral de uma tabela de contingência utilizada na AC simples

| | | Variável B | | | | | | Total |
|------------|-----|------------|----------|-----|----------|-----|----------|----------|
| | | 1 | 2 | ... | j | ... | J | |
| Variável A | 1 | n_{11} | n_{12} | ... | n_{1j} | ... | n_{1J} | $n_{1.}$ |
| | 2 | n_{21} | n_{22} | ... | n_{2j} | ... | n_{2J} | $n_{2.}$ |
| | ... | ... | ... | ... | ... | ... | ... | ... |
| | i | n_{i1} | n_{i2} | ... | n_{ij} | ... | n_{iJ} | $n_{i.}$ |
| | ... | ... | ... | ... | ... | ... | ... | ... |
| | l | n_{l1} | n_{l2} | ... | n_{lj} | ... | n_{lJ} | $n_{l.}$ |
| Total | | $n_{.1}$ | $n_{.2}$ | ... | $n_{.j}$ | ... | $n_{.J}$ | $n_{..}$ |

Em que:

n_{ij} é a frequência observada na i-ésima categoria da variável A e j-ésima categoria da variável B;

$n_{i.}$ é frequência total da i-ésima categoria da variável A, ou seja, o total marginal da i-ésima linha, dado por: $n_{i.} = \sum_{j=1}^J n_{ij}$;

$n_{.j}$ é a frequência total da j-ésima categoria da variável B, ou seja, o total marginal da j-ésima coluna, dado por $n_{.j} = \sum_{i=1}^I n_{ij}$;

$n_{..}$ é o total geral da tabela, dado por $n_{..} = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$.

A AC se inicia com a Tabela 1 representada numa matriz $W_{I \times J}$ denominada matriz de dados.

$$W = \begin{bmatrix} n_{11} & n_{12} & \dots & n_{1j} & \dots & n_{1J} \\ n_{21} & n_{22} & \dots & n_{2j} & \dots & n_{2J} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ n_{i1} & n_{i2} & \dots & n_{ij} & \dots & n_{iJ} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ n_{I1} & n_{I2} & \dots & n_{Ij} & \dots & n_{IJ} \end{bmatrix}$$

No presente texto utilizar-se-á um exemplo ilustrativo com duas variáveis categóricas A e B, com 4 e 3 categorias respectivamente, cujas frequências são apresentadas na tabela de contingência dada a seguir. Este exemplo ilustrativo será retomado toda vez que for apresentado um novo procedimento para execução da AC.

Tabela 2 – Exemplo ilustrativo: tabela de contingência com as frequências observadas das categorias de duas variáveis A e B.

| Variável A | Variável B | | | Total |
|------------|------------|-----|-----|-------|
| | B1 | B2 | B3 | |
| A1 | 50 | 50 | 50 | 150 |
| A2 | 100 | 300 | 100 | 500 |
| A3 | 120 | 0 | 120 | 240 |
| A4 | 100 | 10 | 0 | 110 |
| Total | 370 | 360 | 270 | 1000 |

Fonte: adaptado de Khattree e Naik, 2000, p. 445.

Portanto, a matriz de dados W referente à tabela 2 é dada por,

$$W = \begin{bmatrix} 50 & 50 & 50 \\ 100 & 300 & 100 \\ 120 & 0 & 120 \\ 100 & 10 & 0 \end{bmatrix}$$

Pode-se perceber que a matriz de dados W contém os mesmos valores numéricos da tabela de contingência, exceto pelos totais marginais (total das linhas e das colunas da tabela) e total geral da tabela.

2.2.2 Matriz de correspondência

Como o objetivo final da AC é representar os relacionamentos (correspondências) entre as categorias das variáveis, é necessário que se conheça a distribuição conjunta dessas variáveis. Sendo assim, faz-se necessário transformar a matriz de dados em uma matriz de frequências relativas.

A matriz de frequências relativas é denominada matriz de correspondência P . Esta matriz P não altera as relações proporcionais entre as linhas e as colunas (PAMPLONA, 1998, p. 11; NAITO, 2007, p. 19) e ela é obtida dividindo-se cada termo n_{ij} da matriz W pelo total geral

$$n_{..} = \sum_{i=1}^I \sum_{j=1}^J n_{ij}:$$

$$P = \frac{1}{n_{..}} \cdot W = \begin{bmatrix} \frac{n_{11}}{n_{..}} & \frac{n_{12}}{n_{..}} & \dots & \frac{n_{1j}}{n_{..}} & \dots & \frac{n_{1J}}{n_{..}} \\ \frac{n_{21}}{n_{..}} & \frac{n_{22}}{n_{..}} & \dots & \frac{n_{2j}}{n_{..}} & \dots & \frac{n_{2J}}{n_{..}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{n_{i1}}{n_{..}} & \frac{n_{i2}}{n_{..}} & \dots & \frac{n_{ij}}{n_{..}} & \dots & \frac{n_{iJ}}{n_{..}} \\ \frac{n_{..}}{n_{..}} & \frac{n_{..}}{n_{..}} & \dots & \frac{n_{..}}{n_{..}} & \dots & \frac{n_{..}}{n_{..}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{n_{I1}}{n_{..}} & \frac{n_{I2}}{n_{..}} & \dots & \frac{n_{Ij}}{n_{..}} & \dots & \frac{n_{IJ}}{n_{..}} \\ \frac{n_{..}}{n_{..}} & \frac{n_{..}}{n_{..}} & \dots & \frac{n_{..}}{n_{..}} & \dots & \frac{n_{..}}{n_{..}} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1j} & \dots & p_{1J} \\ p_{21} & p_{22} & \dots & p_{2j} & \dots & p_{2J} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{i1} & p_{i2} & \dots & p_{ij} & \dots & p_{iJ} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{I1} & p_{I2} & \dots & p_{Ij} & \dots & p_{IJ} \end{bmatrix}$$

Portanto, $p_{ij} = \frac{n_{ij}}{n_{..}}$ em que $0 \leq p_{ij} \leq 1 \forall i \text{ e } j$ e $\sum_{i=1}^I \sum_{j=1}^J p_{ij} = 1$.

A matriz de correspondência contém as estimativas de probabilidades conjuntas, isto é, de ocorrência simultânea dos eventos i e j .

A matriz de correspondência do exemplo ilustrativo da Tabela 2 é dada por:

$$P = \frac{1}{1000} \cdot W = \begin{bmatrix} \frac{50}{1000} & \frac{50}{1000} & \frac{50}{1000} \\ \frac{100}{1000} & \frac{300}{1000} & \frac{100}{1000} \\ \frac{120}{1000} & \frac{0}{1000} & \frac{120}{1000} \\ \frac{100}{1000} & \frac{1000}{1000} & \frac{1000}{1000} \\ \frac{100}{1000} & \frac{10}{1000} & \frac{0}{1000} \\ \frac{1000}{1000} & \frac{1000}{1000} & \frac{1000}{1000} \end{bmatrix} = \begin{bmatrix} 0,05 & 0,05 & 0,05 \\ 0,10 & 0,30 & 0,10 \\ 0,12 & 0 & 0,12 \\ 0,10 & 0,01 & 0 \end{bmatrix}.$$

2.2.3 Massa de linha e coluna

São as probabilidades marginais obtidas a partir da matriz P .

$$p_{i\cdot} = \sum_{j=1}^J p_{ij} = \frac{n_{i\cdot}}{n_{\cdot\cdot}} \quad \text{e} \quad \sum_{i=1}^I p_{i\cdot} = 1$$

$$p_{\cdot j} = \sum_{i=1}^I p_{ij} = \frac{n_{\cdot j}}{n_{\cdot\cdot}} \quad \text{e} \quad \sum_{j=1}^J p_{\cdot j} = 1.$$

Em termos práticos representam a importância relativa da i -ésima e j -ésima categorias das variáveis A e B , respectivamente.

Desta forma define-se o vetor massa de linha \mathbf{m}_r que contém o total de cada linha da matriz P , assim como o vetor massa de coluna \mathbf{m}_c contém o total de cada coluna da matriz P :

$$\mathbf{m}_r^t = [p_{1\cdot} \quad p_{2\cdot} \quad \dots \quad p_{i\cdot} \quad \dots \quad p_{I\cdot}]$$

$$\mathbf{m}_c^t = [p_{\cdot 1} \quad p_{\cdot 2} \quad \dots \quad p_{\cdot j} \quad \dots \quad p_{\cdot J}]$$

Em AC estes vetores de massa de linha e coluna são obtidos a partir das seguintes operações matriciais, realizadas diretamente na matriz de correspondência P .

$$\mathbf{m}_r = P_{I \times J} \cdot \mathbf{1}_J = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1j} & \dots & p_{1J} \\ p_{21} & p_{22} & \dots & p_{2j} & \dots & p_{2J} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{i1} & p_{i2} & \dots & p_{ij} & \dots & p_{iJ} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{I1} & p_{I2} & \dots & p_{Ij} & \dots & p_{IJ} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} p_{1\cdot} \\ p_{2\cdot} \\ \vdots \\ p_{i\cdot} \\ \vdots \\ p_{I\cdot} \end{bmatrix}$$

$$\mathbf{m}_c = (P^t)_{J \times I} \cdot \mathbf{1}_I = \begin{bmatrix} p_{11} & p_{21} & \dots & p_{i1} & \dots & p_{I1} \\ p_{12} & p_{22} & \dots & p_{i2} & \dots & p_{I2} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{1j} & p_{2j} & \dots & p_{ij} & \dots & p_{Ij} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{1J} & p_{2J} & \dots & p_{iJ} & \dots & p_{IJ} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} p_{\cdot 1} \\ p_{\cdot 2} \\ \vdots \\ p_{\cdot j} \\ \vdots \\ p_{\cdot J} \end{bmatrix}$$

2.2.4 Perfil de linha e coluna

Inicialmente definem-se as matrizes diagonais dos vetores massa de linha e de coluna:

$$D_r = \text{diag}(m_r) = \begin{bmatrix} p_{1\cdot} & 0 & \dots & 0 & \dots & 0 \\ 0 & p_{2\cdot} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p_{i\cdot} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & p_{I\cdot} \end{bmatrix}$$

$$D_c = \text{diag}(m_c) = \begin{bmatrix} p_{\cdot 1} & 0 & \dots & 0 & \dots & 0 \\ 0 & p_{\cdot 2} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p_{\cdot j} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & p_{\cdot J} \end{bmatrix}$$

A partir das células de uma tabela de frequência, a cada linha i pode-se associar um vetor de probabilidades condicionais $r_i = n_{ij}/n_{i\cdot}$, $j = 1, 2, \dots, J$, assim como, para cada coluna j pode-se associar um vetor de probabilidades condicionais $c_j = n_{ij}/n_{\cdot j}$, $i = 1, 2, \dots, I$. Esses vetores são denominados de perfil de linha e perfil de coluna, respectivamente (BENZÉCRI, 1992, p.9, tradução nossa).

A seguir descrevem-se os cálculos matriciais para obtenção dos perfis de linha e coluna.

Perfil de linha i :

$$r_i^t = \begin{bmatrix} \frac{p_{i1}}{p_{i\cdot}} & \frac{p_{i2}}{p_{i\cdot}} & \dots & \frac{p_{ij}}{p_{i\cdot}} & \dots & \frac{p_{iJ}}{p_{i\cdot}} \end{bmatrix}$$

Perfil de coluna j :

$$c_j^t = \begin{bmatrix} \frac{p_{1j}}{p_{\cdot j}} & \frac{p_{2j}}{p_{\cdot j}} & \dots & \frac{p_{ij}}{p_{\cdot j}} & \dots & \frac{p_{Ij}}{p_{\cdot j}} \end{bmatrix}$$

Cada elemento do vetor r_i^t representa a probabilidade de ocorrer o evento j condicional ao evento i . Interpretação análoga pode ser feita para c_j^t .

Matriz de perfil de linha:

$$R = D_r^{-1}P = \begin{bmatrix} \frac{1}{p_{1\cdot}} & 0 & \dots & 0 & \dots & 0 \\ 0 & \frac{1}{p_{2\cdot}} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \frac{1}{p_{i\cdot}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 0 & \dots & \frac{1}{p_{I\cdot}} \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1j} & \dots & p_{1J} \\ p_{21} & p_{22} & \dots & p_{2j} & \dots & p_{2J} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{i1} & p_{i2} & \dots & p_{ij} & \dots & p_{iJ} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{I1} & p_{I2} & \dots & p_{Ij} & \dots & p_{IJ} \end{bmatrix}$$

$$R = \begin{bmatrix} \frac{p_{11}}{p_{1\cdot}} & \frac{p_{12}}{p_{1\cdot}} & \dots & \frac{p_{1j}}{p_{1\cdot}} & \dots & \frac{p_{1J}}{p_{1\cdot}} \\ \frac{p_{21}}{p_{2\cdot}} & \frac{p_{22}}{p_{2\cdot}} & \dots & \frac{p_{2j}}{p_{2\cdot}} & \dots & \frac{p_{2J}}{p_{2\cdot}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{p_{i1}}{p_{i\cdot}} & \frac{p_{i2}}{p_{i\cdot}} & \dots & \frac{p_{ij}}{p_{i\cdot}} & \dots & \frac{p_{iJ}}{p_{i\cdot}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{p_{I1}}{p_{I\cdot}} & \frac{p_{I2}}{p_{I\cdot}} & \dots & \frac{p_{Ij}}{p_{I\cdot}} & \dots & \frac{p_{IJ}}{p_{I\cdot}} \end{bmatrix}$$

Matriz de perfil de coluna:

$$C = D_c^{-1}P^t = \begin{bmatrix} \frac{1}{p_{\cdot 1}} & 0 & \dots & 0 & \dots & 0 \\ 0 & \frac{1}{p_{\cdot 2}} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \frac{1}{p_{\cdot j}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 0 & \dots & \frac{1}{p_{\cdot J}} \end{bmatrix} \begin{bmatrix} p_{11} & p_{21} & \dots & p_{i1} & \dots & p_{I1} \\ p_{12} & p_{22} & \dots & p_{i2} & \dots & p_{I2} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{1j} & p_{2j} & \dots & p_{ij} & \dots & p_{Ij} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{1J} & p_{2J} & \dots & p_{iJ} & \dots & p_{IJ} \end{bmatrix}$$

$$C = \begin{bmatrix} \frac{p_{11}}{p_{\cdot 1}} & \frac{p_{21}}{p_{\cdot 1}} & \dots & \frac{p_{i1}}{p_{\cdot 1}} & \dots & \frac{p_{I1}}{p_{\cdot 1}} \\ \frac{p_{12}}{p_{\cdot 2}} & \frac{p_{22}}{p_{\cdot 2}} & \dots & \frac{p_{i2}}{p_{\cdot 2}} & \dots & \frac{p_{I2}}{p_{\cdot 2}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{p_{1j}}{p_{\cdot j}} & \frac{p_{2j}}{p_{\cdot j}} & \dots & \frac{p_{ij}}{p_{\cdot j}} & \dots & \frac{p_{Ij}}{p_{\cdot j}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{p_{1J}}{p_{\cdot J}} & \frac{p_{2J}}{p_{\cdot J}} & \dots & \frac{p_{iJ}}{p_{\cdot J}} & \dots & \frac{p_{IJ}}{p_{\cdot J}} \end{bmatrix}$$

Estes conceitos de perfis de linha e coluna são importantes quando se deseja comparar as linhas entre si ou ainda as colunas entre si. Neste contexto, faz-se necessário transformar a matriz P objetivando-se eliminar a influência das suas respectivas marginais (NAITO, 2007, p. 22). O exemplo fictício a seguir ilustra bem esta necessidade.

Supondo que em certo município, foi coletada aleatoriamente uma amostra representativa de 1000 indivíduos cujo objetivo era relacionar o uso de drogas ilícitas com a escolaridade, conforme as tabelas abaixo:

Tabela 3 – Frequências absolutas do uso de drogas ilícitas e escolaridade

| Escolaridade (ensino) | Fez ou faz uso de drogas ilícitas | | Total |
|--------------------------|-----------------------------------|-----|-------|
| | Sim | Não | |
| Fundamental | 75 | 495 | 570 |
| Médio | 63 | 247 | 310 |
| Superior | 42 | 78 | 120 |
| Total | 180 | 820 | 1000 |

Tabela 4 – Frequências relativas do uso de drogas ilícitas e escolaridade

| Escolaridade (ensino) | Fez ou faz uso de drogas ilícitas | | Total |
|--------------------------|-----------------------------------|-------|--------|
| | Sim | Não | |
| Fundamental | 7,5% | 49,5% | 57,0% |
| Médio | 6,3% | 24,7% | 31,0% |
| Superior | 4,2% | 7,8% | 12,0% |
| Total | 18,0% | 82,0% | 100,0% |

Portanto analisando às Tabela 3 e Tabela 4 acima pode-se ter a impressão errônea que a população de nível superior consome menos drogas (4,2%) do que a população de nível fundamental (7,5%). Esta afirmação é correta em termos absolutos, ou seja, sem levar em conta que há mais indivíduos no nível fundamental de educação do que no nível superior.

Para que se possa fazer uma análise, sem que a magnitude (frequências absolutas) das categorias influencie nas comparações, deve-se fazer uma tabela com as “frequências condicionais”, isto é, comparando as frequências relativas com as respectivas frequências marginais.

Tabela 5 – Frequências relativas do uso de drogas ilícitas e escolaridade

| Escolaridade (ensino) | Fez ou faz uso de drogas ilícitas | | Total |
|--------------------------|-----------------------------------|-------|--------|
| | Sim | Não | |
| Fundamental | 13,2% | 86,8% | 100,0% |
| Médio | 20,3% | 79,7% | 100,0% |
| Superior | 35,0% | 65,0% | 100,0% |
| Total | 18,0% | 82,0% | 100,0% |

Na Tabela 5 acima as categorias de escolaridade possuem o mesmo “peso”, e portanto percebe-se claramente que proporcionalmente consome-se mais drogas ilícitas no nível superior do que no ensino fundamental.

É importante ressaltar que os perfis de linha e coluna foram constituídos com o objetivo de se compará-los, porém sem a influência de suas respectivas massas. Surge então um problema, pois cada um desses perfis está intimamente ligado à sua respectiva massa. Adiante no texto será apresentado o conceito de inércia em conexão com este problema.

De volta ao exemplo ilustrativo da Tabela 2, tem-se que:

$$\mathbf{m}_r = P \cdot \mathbf{1}_I = \begin{bmatrix} 0,15 \\ 0,50 \\ 0,24 \\ 0,11 \end{bmatrix} \quad \text{e} \quad \mathbf{m}_c = P^t \cdot \mathbf{1}_J = \begin{bmatrix} 0,37 \\ 0,36 \\ 0,27 \end{bmatrix}$$

$$D_r = \text{diag}(\mathbf{m}_r) = \begin{bmatrix} 0,15 & 0 & 0 & 0 \\ 0 & 0,50 & 0 & 0 \\ 0 & 0 & 0,24 & 0 \\ 0 & 0 & 0 & 0,11 \end{bmatrix}$$

$$D_c = \text{diag}(\mathbf{m}_c) = \begin{bmatrix} 0,37 & 0 & 0 \\ 0 & 0,36 & 0 \\ 0 & 0 & 0,27 \end{bmatrix}$$

Matriz de perfil de linha

$$R = D_r^{-1}P = \begin{bmatrix} 0,33 & 0,33 & 0,33 \\ 0,20 & 0,60 & 0,20 \\ 0,50 & 0 & 0,50 \\ 0,91 & 0,09 & 0 \end{bmatrix},$$

formada pelos vetores:

$$\mathbf{r}_1^t = [0,33 \quad 0,33 \quad 0,33]$$

$$\mathbf{r}_2^t = [0,20 \quad 0,60 \quad 0,20]$$

$$\mathbf{r}_3^t = [0,50 \quad 0,00 \quad 0,50]$$

$$\mathbf{r}_4^t = [0,91 \quad 0,09 \quad 0,00]$$

Matriz de perfil de coluna

$$C = D_c^{-1}P^t = \begin{bmatrix} 0,14 & 0,27 & 0,32 & 0,27 \\ 0,14 & 0,83 & 0 & 0,03 \\ 0,19 & 0,37 & 0,44 & 0 \end{bmatrix},$$

formada pelos vetores:

$$\mathbf{c}_1^t = [0,14 \quad 0,27 \quad 0,32 \quad 0,27]$$

$$\mathbf{c}_2^t = [0,14 \quad 0,83 \quad 0,00 \quad 0,03]$$

$$\mathbf{c}_3^t = [0,19 \quad 0,37 \quad 0,44 \quad 0,00]$$

2.2.5 A nuvem de pontos

Como a ideia básica da AC é a representação gráfica das relações (correspondências) entre as categorias das variáveis envolvidas, faz-se necessário representar os perfis de linha e coluna, ou seja, as categorias de linha e coluna em um mesmo espaço multidimensional. Assim, os perfis de linha e coluna podem ser representados em um espaço vetorial J -dimensional, \mathbb{R}^J , e I -dimensional, \mathbb{R}^I , respectivamente (NAITO, 2007, p. 25).

Cada um desses espaços vetoriais formará uma nuvem $N(I)$ e $N(J)$ com I e J pontos, respectivamente, formados pelas componentes de cada um dos vetores dos perfis. Essas componentes serão consideradas as coordenadas de cada ponto vetorial (perfil).

Admite-se que as componentes de cada vetor seguem uma distribuição multinomial condicional ao total da linha ou coluna e obviamente sua soma é 1, o que gera uma dependência linear entre tais componentes.

Assim sendo, cada um desses pontos vetoriais (perfil de linha ou coluna) podem ser representados em uma dimensão menor do que originalmente foram projetados, ou seja, dimensão $(J - 1)$ para representar os perfis de linha e dimensão $(I - 1)$ para representar os perfis de coluna. Retomando ao exemplo ilustrativo da Tabela 2, tem-se que cada um dos 4 vetores que formam o perfil de linha possuem 3 componentes, podendo então representá-los em um espaço euclidiano tridimensional.

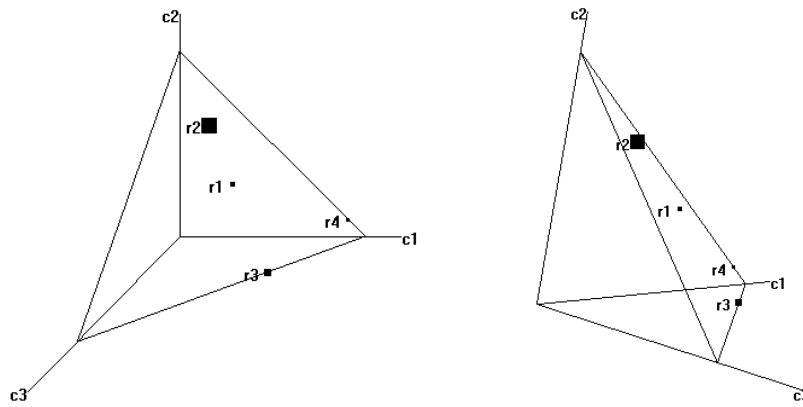


Figura 19 – Representação tri-dimensional dos perfis de linha sob dois ângulos de visão.

Pode-se perceber que estes pontos vetoriais (perfis de linha) foram originalmente representados em um espaço tridimensional, porém é possível perceber que todos os pontos estão sobre o mesmo plano (uma dimensão menor), denominado de simplex, devido a dependência linear entre suas componentes, que somam 1.

2.2.6 Centro de gravidade ou centroide

Segundo Benzécri (1992, p. 28) o centro de gravidade de um sistema de pontos com massas que lhe são atribuídas (números positivos ou zeros) é uma generalização espacial da média aritmética. Sendo assim, para se determinar o centro de gravidade ou centroide, deve-se fazer a média das componentes (coordenadas) referentes a cada eixo de base, ponderadas pela sua respectiva massa.

Considerando-se a nuvem $N(I)$, para se determinar a j -ésima componente (coordenada) referente ao centroide deve-se calcular a média entre as j -ésimas componentes dos perfis ponderando-as pelas suas respectivas massas, conforme apresentado a seguir:

$$g_r = \frac{\sum_{i=1}^I \frac{p_{ij}}{p_{i.}} \cdot m_i}{\sum_{i=1}^I m_i} = \frac{\sum_{i=1}^I \frac{p_{ij}}{p_{i.}} \cdot m_i}{1} = \sum_{i=1}^I \frac{p_{ij}}{p_{i.}} \cdot p_{i.} = p_{.j}$$

Conforme apresentado, o centroide ou perfil médio de linha pode ser determinado pelo próprio vetor de massa de coluna e de forma similar o perfil médio de coluna da nuvem N(J) pode ser determinado pelo próprio vetor de massa de linha. Aplicando-se ao exemplo ilustrativo da Tabela 2, tem-se:

Os vetores dos perfis de linha são:

$$\mathbf{r}_1^t = [0,33 \quad 0,33 \quad 0,33]$$

$$\mathbf{r}_2^t = [0,20 \quad 0,60 \quad 0,20]$$

$$\mathbf{r}_3^t = [0,50 \quad 0,00 \quad 0,50]$$

$$\mathbf{r}_4^t = [0,91 \quad 0,09 \quad 0,00]$$

E o vetor de massa de linha é:

$$\mathbf{m}_r = \begin{bmatrix} 0,15 \\ 0,50 \\ 0,24 \\ 0,11 \end{bmatrix}$$

A média ponderada da primeira componente do centroide é dada por:

$$g_1 = \frac{0,33 \cdot 0,15 + 0,2 \cdot 0,5 + 0,5 \cdot 0,24 + 0,91 \cdot 0,11}{0,15 + 0,5 + 0,24 + 0,11} = 0,37$$

Procede-se da mesma forma para determinar-se as outras componentes do centroide, ou de forma matricial, calcula-se o perfil médio da linha como:

$$\mathbf{g}_r = \mathbf{R}^t \cdot \mathbf{m}_r = \begin{bmatrix} 0,33 & 0,20 & 0,50 & 0,91 \\ 0,33 & 0,60 & 0 & 0,09 \\ 0,33 & 0,20 & 0,50 & 0 \end{bmatrix} \begin{bmatrix} 0,15 \\ 0,50 \\ 0,24 \\ 0,11 \end{bmatrix}$$

$$\mathbf{g}_r^t = [0,37 \quad 0,36 \quad 0,27]$$

Procedendo-se da mesma forma encontra-se o perfil médio da coluna:

$$\mathbf{g}_c = \mathbf{C}^t \cdot \mathbf{m}_c = \begin{bmatrix} 0,14 & 0,14 & 0,19 \\ 0,27 & 0,83 & 0,37 \\ 0,32 & 0 & 0,44 \\ 0,27 & 0,03 & 0 \end{bmatrix} \begin{bmatrix} 0,37 \\ 0,36 \\ 0,27 \end{bmatrix}$$

$$\mathbf{g}_c^t = [0,15 \quad 0,50 \quad 0,24 \quad 0,11]$$

Projetando-se o perfil médio da linha (vermelho) juntamente com os perfis de linha, obtém-se a nuvem de pontos N(l):

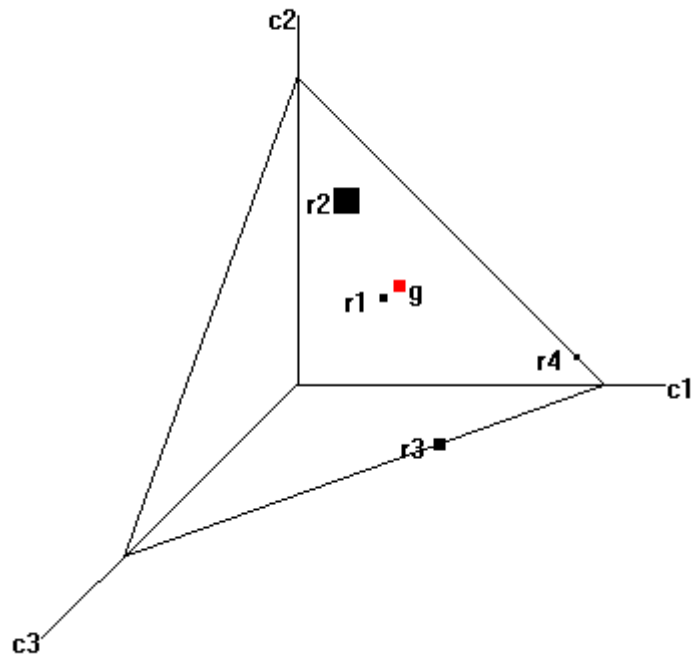


Figura 20 – Representação do perfil médio de linha (g) com os perfis de linha no espaço tridimensional

2.2.7 Métrica euclidiana ponderada ou distância de qui-quadrado

Uma menor ou maior distância entre os pontos de uma nuvem pode indicar, respectivamente, relacionamento ou não entre as categorias das variáveis. Assim sendo, faz-se necessário definir uma métrica para medir tal distância. Usualmente utiliza-se a distância euclidiana, a qual tem como base o Teorema de Pitágoras para definir a distância entre dois pontos.

Segundo Souza (2004, p. 78), define-se a distância euclidiana entre dois vetores $\mathbf{r}_1^t = [x_1 \ y_1 \ z_1]$ e $\mathbf{r}_2^t = [x_2 \ y_2 \ z_2]$ do \mathbb{R}^3 como:

$$d(\mathbf{r}_1, \mathbf{r}_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

Logo, a distância quadrática entre estes pontos é

$$d^2(\mathbf{r}_1, \mathbf{r}_2) = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2.$$

O centroide da nuvem representa o ponto que é esperado se as variáveis da linha e da coluna da tabela de contingência forem independentes. Portanto, medir a distância que cada ponto (perfil de linha ou coluna) possui de seu respectivo centroide significa quantificar a dispersão. Sendo assim, a distância euclidiana de um ponto (perfil de linha) qualquer ao seu centroide na nuvem $N(I)$ no espaço vetorial R^J é dada por:

$$d^2(\mathbf{r}_i, \mathbf{g}_r) = \left(\frac{p_{i1}}{p_{i\cdot}} - p_{\cdot 1}\right)^2 + \left(\frac{p_{i2}}{p_{i\cdot}} - p_{\cdot 2}\right)^2 + \dots + \left(\frac{p_{iJ}}{p_{i\cdot}} - p_{\cdot J}\right)^2.$$

Semelhantemente, em relação à nuvem $N(J)$ tem-se:

$$d^2(\mathbf{c}_i, \mathbf{g}_r) = \left(\frac{p_{1j}}{p_{\cdot j}} - p_{1\cdot}\right)^2 + \left(\frac{p_{2j}}{p_{\cdot j}} - p_{2\cdot}\right)^2 + \dots + \left(\frac{p_{Ij}}{p_{\cdot j}} - p_{I\cdot}\right)^2.$$

A distância euclidiana é uma medida absoluta da dispersão de uma coordenada de um perfil em relação à respectiva coordenada do centroide, sem levar em consideração a dispersão relativa que essa distância representa. Para um melhor entendimento desta afirmação, considere o seguinte exemplo:

Seja um vetor (perfil de linha) com as coordenadas $\mathbf{r} = [0,05 \ 0,6 \ 0,35]$ e seu centroide com as coordenadas $\mathbf{g} = [0,1 \ 0,65 \ 0,25]$.

Ao medir a distância entre os vetores \mathbf{r} e \mathbf{g} pela distância euclidiana tem-se que:

$$d^2(\mathbf{r}, \mathbf{g}) = (0,05 - 0,1)^2 + (0,60 - 0,65)^2 + (0,35 - 0,25)^2$$

$$d^2(\mathbf{r}, \mathbf{g}) = (0,05)^2 + (0,05)^2 + (0,1)^2$$

Observa-se pela distância euclidiana que a diferença das componentes da primeira e da segunda coordenadas são iguais, porém não em termos relativos à magnitude dos valores das coordenadas, visto que 0,05 em relação ao valor 0,1 é relativamente maior do que em relação ao valor 0,65. Sendo assim, pondera-se a distancia pelo inverso da componente referente ao centroide.

$$d^2(\mathbf{r}, \mathbf{g}) = \frac{1}{0,1} (0,05 - 0,1)^2 + \frac{1}{0,65} (0,60 - 0,65)^2 + \frac{1}{0,25} (0,35 - 0,25)^2$$

$$d^2(\mathbf{r}, \mathbf{g}) = 10(0,05)^2 + 1,538(0,05)^2 + 4(0,1)^2$$

Portanto, a métrica entre os pontos da nuvem e o seu respectivo centro de gravidade é dada pela distância euclidiana ponderada, também denominada distância de qui-quadrado (PAMPLONA, 1998, p. 16), por sua semelhança com a estatística do teste de qui-quadrado.

$$d^2(\mathbf{r}_i, \mathbf{g}_r) = \frac{1}{p_{\cdot 1}} \left(\frac{p_{i1}}{p_{i\cdot}} - p_{\cdot 1} \right)^2 + \frac{1}{p_{\cdot 2}} \left(\frac{p_{i2}}{p_{i\cdot}} - p_{\cdot 2} \right)^2 + \dots + \frac{1}{p_{\cdot J}} \left(\frac{p_{iJ}}{p_{i\cdot}} - p_{\cdot J} \right)^2$$

$$d^2(\mathbf{r}_i, \mathbf{g}_r) = \sum_{j=1}^J \frac{\left(\frac{p_{ij}}{p_{i\cdot}} - p_{\cdot j} \right)^2}{p_{\cdot j}}$$

Em notação matricial, tem-se:

$$d^2(\mathbf{r}_i, \mathbf{g}_r) = (\mathbf{r}_i - \mathbf{g}_r)^t D_c^{-1} (\mathbf{r}_i - \mathbf{g}_r)$$

Pode-se também definir a distância entre dois perfis de linha quaisquer por:

$$d^2(\mathbf{r}_i, \mathbf{r}_{i'}) = \frac{1}{p_{\cdot 1}} \left(\frac{p_{i1}}{p_{i\cdot}} - \frac{p_{i'1}}{p_{i'\cdot}} \right)^2 + \frac{1}{p_{\cdot 2}} \left(\frac{p_{i2}}{p_{i\cdot}} - \frac{p_{i'2}}{p_{i'\cdot}} \right)^2 + \dots + \frac{1}{p_{\cdot J}} \left(\frac{p_{iJ}}{p_{i\cdot}} - \frac{p_{i'J}}{p_{i'\cdot}} \right)^2$$

$$d^2(\mathbf{r}_i, \mathbf{r}_{i'}) = \sum_{j=1}^J \frac{\left(\frac{p_{ij}}{p_{i\cdot}} - \frac{p_{i'j}}{p_{i'\cdot}} \right)^2}{p_{\cdot j}}$$

Em notação matricial, tem-se:

$$d^2(\mathbf{r}_i, \mathbf{r}_{i'}) = (\mathbf{r}_i - \mathbf{r}_{i'})^t D_c^{-1} (\mathbf{r}_i - \mathbf{r}_{i'})$$

De forma semelhante pode-se definir a distância entre dois perfis coluna por:

$$d^2(\mathbf{c}_j, \mathbf{c}_{j'}) = \sum_{i=1}^I \frac{\left(\frac{p_{ij}}{p_{.j}} - \frac{p_{ij'}}{p_{.j'}} \right)^2}{p_i}$$

Em notação matricial, tem-se:

$$d^2(\mathbf{c}_j, \mathbf{c}_{j'}) = (\mathbf{c}_j - \mathbf{c}_{j'})^t D_r^{-1} (\mathbf{c}_j - \mathbf{c}_{j'})$$

Outra justificativa para o uso da distância de qui-quadrado como métrica entre as distâncias dos pontos é que a mesma satisfaz a propriedade da equivalência distribucional (GREENACRE, 1984, p.95). Este princípio permite que a inércia, assunto apresentado na próxima seção, da nuvem de linha N(I) seja a mesma da nuvem de coluna N(J).

O princípio da equivalência distribucional permite que dois perfis idênticos (ou equivalentes distribucionalmente), possam ter suas frequências absolutas somadas para gerar uma única linha. Esta nova linha terá o mesmo perfil, sendo a sua massa igual à soma das massas das linhas somadas, sem que isso altere a distância entre os perfis da outra nuvem (Jelihovschi e Ferraz, p. 121).

O exemplo descrito a seguir mostra a propriedade da equivalência distribucional descrita anteriormente.

Considere uma matriz de correspondência $G_{3 \times 3}$ em que a linha 3 é k vezes a linha 2.

$$G = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ kp_{21} & kp_{22} & kp_{23} \end{bmatrix}$$

Matriz de perfil de linha:

$$R = \begin{bmatrix} r_1^t \\ r_2^t \\ r_3^t \end{bmatrix} = \begin{bmatrix} \frac{p_{11}}{p_{1\cdot}} & \frac{p_{12}}{p_{1\cdot}} & \frac{p_{13}}{p_{1\cdot}} \\ \frac{p_{21}}{p_{2\cdot}} & \frac{p_{22}}{p_{2\cdot}} & \frac{p_{23}}{p_{2\cdot}} \\ \frac{kp_{21}}{kp_{2\cdot}} & \frac{kp_{22}}{kp_{2\cdot}} & \frac{kp_{23}}{kp_{2\cdot}} \end{bmatrix}$$

Matriz de perfil de coluna:

$$C = \begin{bmatrix} c_1^t \\ c_2^t \\ c_3^t \end{bmatrix} = \begin{bmatrix} \frac{p_{11}}{p_{\cdot 1}} & \frac{p_{21}}{p_{\cdot 1}} & \frac{kp_{21}}{p_{\cdot 1}} \\ \frac{p_{12}}{p_{\cdot 2}} & \frac{p_{22}}{p_{\cdot 2}} & \frac{kp_{22}}{p_{\cdot 2}} \\ \frac{p_{13}}{p_{\cdot 3}} & \frac{p_{23}}{p_{\cdot 3}} & \frac{kp_{23}}{p_{\cdot 3}} \end{bmatrix}$$

Que nos leva a verificar que o segundo e o terceiro perfis de linha são idênticos, sendo projetados no mesmo ponto da nuvem $N(l)$. Assim, pode-se somar a segunda e terceira linhas da matriz de correspondência G , gerando-se assim uma nova matriz G' e por consequência novos perfis de linha e coluna.

$$G' = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ (k+1)p_{21} & (k+1)p_{22} & (k+1)p_{23} \end{bmatrix}$$

novos perfis de linha

$$R' = \begin{bmatrix} r_1^t \\ r_0^t \end{bmatrix} = \begin{bmatrix} \frac{p_{11}}{p_{1\cdot}} & \frac{p_{12}}{p_{1\cdot}} & \frac{p_{13}}{p_{1\cdot}} \\ \frac{(k+1)p_{21}}{(k+1)p_{2\cdot}} & \frac{(k+1)p_{22}}{(k+1)p_{2\cdot}} & \frac{(k+1)p_{23}}{(k+1)p_{2\cdot}} \end{bmatrix}$$

Sendo assim, percebe-se que o vetor r_0' é idêntico aos vetores de origem r_2 e r_3 . Portanto, o importante resultado oriundo do princípio da equivalência distribucional é que depois de unificados os perfis semelhantes, sua projeção não se altera na nuvem de origem, como também não modifica as distâncias da outra nuvem, conforme exemplificado a seguir.

Matriz de perfil de coluna:

$$C' = \begin{bmatrix} c_1^t \\ c_2^t \\ c_3^t \end{bmatrix} = \begin{bmatrix} \frac{p_{11}}{p_{\cdot 1}} & \frac{(k+1)p_{21}}{p_{\cdot 1}} \\ \frac{p_{12}}{p_{\cdot 2}} & \frac{(k+1)p_{22}}{p_{\cdot 2}} \\ \frac{p_{13}}{p_{\cdot 3}} & \frac{(k+1)p_{23}}{p_{\cdot 3}} \end{bmatrix}$$

Calculando a distância entre o ponto correspondente ao vetor c_1^t (antes da união dos vetores semelhantes) e o seu respectivo centro de gravidade g_c , através da distância de qui-quadrado, tem-se:

$$d^2(c_1, g_c) = \frac{1}{p_{1.}} \left(\frac{p_{11}}{p_{.1}} - p_{1.} \right)^2 + \frac{1}{p_{2.}} \left(\frac{p_{21}}{p_{.1}} - p_{2.} \right)^2 + \frac{1}{kp_{2.}} \left(\frac{kp_{21}}{p_{.1}} - kp_{2.} \right)^2$$

$$d^2(c_1, g_c) = \frac{1}{p_{1.}} \left(\frac{p_{11}}{p_{.1}} - p_{1.} \right)^2 + \frac{1}{p_{2.}} \left(\frac{p_{21}}{p_{.1}} - p_{2.} \right)^2 + \frac{k^2}{kp_{2.}} \left(\frac{p_{21}}{p_{.1}} - p_{2.} \right)^2$$

$$d^2(c_1, g_c) = \frac{1}{p_{1.}} \left(\frac{p_{11}}{p_{.1}} - p_{1.} \right)^2 + \frac{(k+1)}{p_{2.}} \left(\frac{p_{21}}{p_{.1}} - p_{2.} \right)^2$$

$$d^2(c_1, g_c) = \frac{1}{p_{1.}} \left(\frac{p_{11}}{p_{.1}} - p_{1.} \right)^2 + \frac{(k+1)^2}{(k+1)p_{2.}} \left(\frac{p_{21}}{p_{.1}} - p_{2.} \right)^2$$

$$d^2(c_1, g_c) = \frac{1}{p_{1.}} \left(\frac{p_{11}}{p_{.1}} - p_{1.} \right)^2 + \frac{1}{(k+1)p_{2.}} \left(\frac{(k+1)p_{21}}{p_{.1}} - (k+1)p_{2.} \right)^2$$

Calculando-se a distância entre c'_1 e o centro de gravidade g'_c

$$d^2(c'_1, g'_c) = \frac{1}{p_{1.}} \left(\frac{p_{11}}{p_{.1}} - p_{1.} \right)^2 + \frac{1}{(k+1)p_{2.}} \left(\frac{(k+1)p_{21}}{p_{.1}} - (k+1)p_{2.} \right)^2$$

Portanto, pode-se observar que $d^2(c_1, g_c)$ é igual $d^2(c'_1, g'_c)$, sendo assim conclui-se que ao utilizar a distância de qui-quadrado para medir a distância entre pontos de uma nuvem, conservam-se as distâncias entre os pontos, mesmo quando se unificam perfis.

Retornando ao exemplo ilustrativo da tabela 2, pode-se calcular a distância entre o quarto perfil de linha e o centroide como:

$$d^2(r_4, g_r) = \frac{1}{p_{.1}} \left(\frac{p_{41}}{p_{4.}} - p_{.1} \right)^2 + \frac{1}{p_{.2}} \left(\frac{p_{42}}{p_{4.}} - p_{.2} \right)^2 + \frac{1}{p_{.3}} \left(\frac{p_{43}}{p_{4.}} - p_{.3} \right)^2$$

$$d^2(r_4, g_r) = \frac{1}{0,37} (0,91 - 0,37)^2 + \frac{1}{0,36} (0,09 - 0,36)^2 + \frac{1}{0,27} (0 - 0,27)^2$$

$$d^2(r_4, g_r) = 1,256$$

Ou de forma matricial

$$d^2(r_4, g_r) = (r_4 - g_r)^t D_c^{-1} (r_4 - g_r)$$

$$d^2(\mathbf{r}_4, \mathbf{g}_r) = [0,539 \quad -0,269 \quad -0,27] \begin{bmatrix} \frac{1}{0,37} & 0 & 0 \\ 0 & \frac{1}{0,36} & 0 \\ 0 & 0 & \frac{1}{0,27} \end{bmatrix} \begin{bmatrix} 0,539 \\ -0,269 \\ -0,27 \end{bmatrix}$$

$$d^2(\mathbf{r}_4, \mathbf{g}_r) = 1,256$$

2.2.8 Inércia

Até este ponto no texto demonstrou-se que os perfis de linha e coluna podem ser representados através de uma nuvem de pontos e que a distância entre cada um desses pontos pode ser medida pela distância euclidiana ponderada ou distância de qui-quadrado.

Com o intuito de medir a dispersão desses pontos em relação ao centro de gravidade, calcula-se a inércia, que é um conceito da física, semelhante ao conceito estatístico de variância. Esta semelhança entre variância e inércia é apresentada por DANTAS (2004, p. 92) que diz:

A variância fornece uma medida de dispersão dos valores da variável aleatória em relação à sua esperança. Ela, a exemplo da esperança, também pode ser interpretada fisicamente. Representando-se a distribuição de probabilidades por meio de uma distribuição de massas, colocando-se em cada valor da variável uma massa igual a sua probabilidade, então o valor da variância coincide com o valor do momento de inércia dessa distribuição de massas em torno de seu centro de gravidade.

Os perfis de linha e coluna seguem uma distribuição multinomial condicionada ao total da linha e da coluna. Então a variância total da nuvem de pontos $N(I)$ pode ser calculada de forma discreta. Sendo assim:

$$\sigma_I^2 = \sum_{i=1}^I [x_i - E(X)]^2 \cdot P(X = x_i)$$

A diferença entre x_i e $E(X)$ é medida pela distância de qui-quadrado já apresentada, ponderada por $P(X = x_i)$, ou seja, sua respectiva massa. Então pode-se reescrever a variância descrita acima como:

$$\hat{\sigma}_I^2 = \sum_{i=1}^I d(r_i, g_r)^2 \cdot m_i$$

Como a distância entre um perfil i e o centroide da nuvem $N(I)$ é dado por $d^2(r_i, g_r) = \sum_{j=1}^J \frac{(p_{ij} - p_{\cdot j})^2}{p_{\cdot j}}$, e m_i corresponde à massa da linha i , ou seja, $p_{i\cdot}$, então a variância ou inércia total da nuvem $N(I)$ é descrita como:

$$\hat{\sigma}_I^2 = In(I) = \sum_{i=1}^I p_{i\cdot} \sum_{j=1}^J \frac{(p_{ij} - p_{\cdot j})^2}{p_{\cdot j}}$$

Que em notação matricial é dada por:

$$In(I) = \text{traço}[D_r(R - \mathbf{1}m_c^t)D_c^{-1}(R - \mathbf{1}m_c^t)^t]$$

De forma semelhante, a inércia total da nuvem J é dada por:

$$\hat{\sigma}_J^2 = In(J) = \sum_{j=1}^J p_{\cdot j} \sum_{i=1}^I \frac{(p_{ij} - p_{i\cdot})^2}{p_{i\cdot}}$$

Que em notação matricial é dada por:

$$In(J) = \text{traço}[D_c(C - \mathbf{1}m_r^t)D_r^{-1}(C - \mathbf{1}m_r^t)^t]$$

Uma vez que a soma das massas (ponderações) vale 1, também pode-se expressar inércia como sendo a média ponderada do quadrado das distâncias de qui-quadrado entre os perfis e o centroide (GREENACRE, 2007, p. 29).

Desenvolvendo-se as inércias de ambas as nuvens descritas anteriormente, pode-se perceber que a utilização da distância de qui-quadrado como métrica entre dois pontos da nuvem preserva o princípio da equivalência distribucional e resulta na igualdade das inércias totais de ambas as nuvens $N(I)$ e $N(J)$, como demonstrado a seguir:

$$\begin{aligned} \hat{\sigma}_I^2 = In(I) &= \sum_{i=1}^I p_{i\cdot} \sum_{j=1}^J \frac{\left(\frac{p_{ij}}{p_{i\cdot}} - p_{\cdot j}\right)^2}{p_{\cdot j}} & \hat{\sigma}_J^2 = In(J) &= \sum_{j=1}^J p_{\cdot j} \sum_{i=1}^I \frac{\left(\frac{p_{ij}}{p_{\cdot j}} - p_{i\cdot}\right)^2}{p_{i\cdot}} \\ \hat{\sigma}_I^2 = In(I) &= \sum_{i=1}^I \sum_{j=1}^J \frac{p_{i\cdot} \left(\frac{p_{ij}}{p_{i\cdot}} - p_{\cdot j}\right)^2}{p_{\cdot j}} & \hat{\sigma}_J^2 = In(J) &= \sum_{j=1}^J \sum_{i=1}^I \frac{p_{\cdot j} \left(\frac{p_{ij}}{p_{\cdot j}} - p_{i\cdot}\right)^2}{p_{i\cdot}} \\ \hat{\sigma}_I^2 = In(I) &= \sum_{i=1}^I \sum_{j=1}^J \frac{p_{i\cdot}^2 \left(\frac{p_{ij}}{p_{i\cdot}} - p_{\cdot j}\right)^2}{p_{i\cdot} p_{\cdot j}} & \hat{\sigma}_J^2 = In(J) &= \sum_{j=1}^J \sum_{i=1}^I \frac{p_{\cdot j}^2 \left(\frac{p_{ij}}{p_{\cdot j}} - p_{i\cdot}\right)^2}{p_{i\cdot} p_{\cdot j}} \\ \hat{\sigma}_I^2 = In(I) &= \sum_{i=1}^I \sum_{j=1}^J \frac{\left(\frac{p_{i\cdot} p_{ij}}{p_{i\cdot}} - p_{i\cdot} p_{\cdot j}\right)^2}{p_{i\cdot} p_{\cdot j}} & \hat{\sigma}_J^2 = In(J) &= \sum_{j=1}^J \sum_{i=1}^I \frac{\left(\frac{p_{\cdot j} p_{ij}}{p_{\cdot j}} - p_{i\cdot} p_{\cdot j}\right)^2}{p_{i\cdot} p_{\cdot j}} \\ \hat{\sigma}_I^2 = In(I) &= \sum_{i=1}^I \sum_{j=1}^J \frac{\left(p_{ij} - p_{i\cdot} p_{\cdot j}\right)^2}{p_{i\cdot} p_{\cdot j}} & \hat{\sigma}_J^2 = In(J) &= \sum_{j=1}^J \sum_{i=1}^I \frac{\left(p_{ij} - p_{i\cdot} p_{\cdot j}\right)^2}{p_{i\cdot} p_{\cdot j}} \end{aligned}$$

Logo,

$$In(I) = In(J) = \phi^2 = \sigma^2$$

Para o exemplo da Tabela 2, ilustra-se a seguir apenas o cálculo da inércia do quarto perfil de linha (r_4) ao centroide,

$$\begin{aligned} \hat{\sigma}_{r_i}^2 &= p_{i\cdot} \sum_{j=1}^J \frac{\left(\frac{p_{ij}}{p_{i\cdot}} - p_{\cdot j}\right)^2}{p_{\cdot j}} \\ \hat{\sigma}_{r_4}^2 &= 0,11 \left[\frac{(0,91 - 0,37)^2}{0,37} + \frac{(0,09 - 0,36)^2}{0,36} + \frac{(0 - 0,27)^2}{0,27} \right] = 0,1382 \end{aligned}$$

Procede-se de forma semelhante para se determinar as inércias dos outros perfis de linha ($\hat{\sigma}_{r_1}^2$, $\hat{\sigma}_{r_2}^2$ e $\hat{\sigma}_{r_3}^2$). Portanto, a inércia total da nuvem $N(I)$ é dada por,

$$\begin{aligned} In(I) = \phi^2 = \hat{\sigma}_I^2 &= \hat{\sigma}_{r_1}^2 + \hat{\sigma}_{r_2}^2 + \hat{\sigma}_{r_3}^2 + \hat{\sigma}_{r_4}^2 \\ \hat{\sigma}_I^2 &= 0,0031 + 0,1281 + 0,1444 + 0,1382 \\ \hat{\sigma}_I^2 &= 0,4138 \end{aligned}$$

Novamente, para o exemplo da tabela 2, ilustra-se a seguir apenas o cálculo da inércia do terceiro perfil de coluna (c_3) ao centroide,

$$\hat{\sigma}_{c_i}^2 = p_{.j} \sum_{i=1}^I \frac{\left(\frac{p_{ij}}{p_{.j}} - p_{.i}\right)^2}{p_{.i}}$$

$$\hat{\sigma}_{c_3}^2 = 0,27 \left[\frac{(0,19 - 0,15)^2}{0,15} + \frac{(0,37 - 0,5)^2}{0,5} + \frac{(0,44 - 0,24)^2}{0,24} + \frac{(0 - 0,11)^2}{0,11} \right]$$

$$\hat{\sigma}_{c_3}^2 = 0,0882$$

Procede-se de forma semelhante para se determinar as inércias dos outros perfis de coluna ($\hat{\sigma}_{c_1}^2$ e $\hat{\sigma}_{c_2}^2$). Portanto, a inércia total da nuvem N(J) é dada por:

$$In(J) = \phi^2 = \hat{\sigma}_J^2 = \hat{\sigma}_{c_1}^2 + \hat{\sigma}_{c_2}^2 + \hat{\sigma}_{c_3}^2$$

$$\hat{\sigma}_J^2 = 0,1369 + 0,1889 + 0,0880$$

$$\hat{\sigma}_J^2 = 0,4138$$

Logo,

$$\phi^2 = \hat{\sigma}_I^2 = \hat{\sigma}_J^2 = 0,4138$$

Os cálculos anteriores descritos sob forma matricial são:

i) a inércia total da nuvem I:

$$\hat{\sigma}_I^2 = \text{traço}[D_r(R - \mathbf{1m}_c^t)D_c^{-1}(R - \mathbf{1m}_c^t)^t]$$

$$\hat{\sigma}_I^2 = \text{traço} \begin{bmatrix} 0,0031 & -0,0026 & 0,0102 & 0,0142 \\ -0,0087 & 0,1281 & -0,1797 & -0,1785 \\ 0,0163 & -0,0862 & 0,1444 & 0,0548 \\ -0,0106 & -0,0393 & 0,0251 & 0,1382 \end{bmatrix} = 0,4138$$

ii) a inércia total da nuvem J:

$$\hat{\sigma}_J^2 = \text{traço}[D_c(C - \mathbf{1m}_r^t)D_r^{-1}(C - \mathbf{1m}_r^t)^t]$$

$$\hat{\sigma}_J^2 = \text{traço} \begin{bmatrix} 0,1369 & -0,1318 & -0,0120 \\ -0,1282 & 0,1889 & -0,0760 \\ -0,0087 & -0,0570 & 0,0880 \end{bmatrix} = 0,4138$$

2.2.9 Relação entre inércia e a estatística do teste de qui-quadrado

O teste de qui-quadrado proposto por Pearson é comumente utilizado para se testar uma hipótese de nulidade que afirma serem independentes duas variáveis categóricas A e B, cujas frequências são apresentadas em uma tabela de contingência.

A estatística do teste de qui-quadrado aplicado a esta tabela de contingência pode ser escrita como (BEH, 2004, p. 261-262):

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - \frac{n_{i.}n_{.j}}{N})^2}{\frac{n_{i.}n_{.j}}{N}}$$

Em que:

- n_{ij} é o valor observado
- $\frac{n_{i.}n_{.j}}{N}$ é o valor esperado para situação de total independência entre as categorias i e j

Ao multiplicar e dividir cada termo por N, pode-se obter uma relação conveniente na AC, entre a estatística do teste de qui-quadrado (χ^2) e a inércia (ϕ^2) descrita anteriormente (KHATTREE e NAIK, 2000, p. 446).

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(\frac{Nn_{ij}}{N} - \frac{Nn_{i.}n_{.j}}{NN})^2}{\frac{Nn_{i.}n_{.j}}{NN}}$$

Como as frequências relativas são dadas por

$$\frac{n_{ij}}{N} = p_{ij}; \quad \frac{n_{i.}}{N} = p_{i.}; \quad \frac{n_{.j}}{N} = p_{.j},$$

tem-se que:

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(Np_{ij} - Np_{i.}p_{.j})^2}{Np_{i.}p_{.j}} = N \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - p_{i.}p_{.j})^2}{p_{i.}p_{.j}}$$

Portanto,

$$\chi^2 = N\phi^2 \quad \text{ou} \quad \phi^2 = \frac{\chi^2}{N}$$

2.2.10 Espaço transformado centrado no centroide

A distância euclidiana, por ser a mais facilmente interpretada, é a métrica mais comumente utilizada para medir a distância entre dois pontos no espaço. Ela permite estabelecer similaridades e diferenças de acordo com a proximidade entre os pontos. Porém, conforme já justificado anteriormente no texto, em AC é utilizada a distância de qui-quadrado para medir a distância entre pontos quaisquer de uma nuvem.

Com o objetivo de equiparar a distância de qui-quadrado utilizada na AC com a distância euclidiana comum na maioria das soluções de redução de dimensionalidade, faz-se necessário ponderar as escalas dos eixos e as coordenadas de todos os pontos das nuvens de linha $N(I)$ e de coluna $N(J)$ por $\frac{1}{\sqrt{p_{\cdot j}}}$ e $\frac{1}{\sqrt{p_{i \cdot}}}$, respectivamente (CZERMAINSKI, 2004, p. 7). Esta ponderação torna a distância euclidiana igual à distância de qui-quadrado, conforme é mostrado a seguir.

A distância entre um perfil de linha r_i ao centroide é:

$$d^2(\mathbf{r}_i, \mathbf{g}_r) = \sum_{j=1}^J \frac{\left(\frac{p_{ij}}{p_{i \cdot}} - p_{\cdot j}\right)^2}{p_{\cdot j}} = \sum_{j=1}^J \frac{1}{\sqrt{p_{\cdot j}}^2} \left(\frac{p_{ij}}{p_{i \cdot}} - p_{\cdot j}\right)^2 = \sum_{j=1}^J \left(\frac{p_{ij}}{\sqrt{p_{\cdot j} p_{i \cdot}}} - \frac{p_{\cdot j}}{\sqrt{p_{\cdot j}}}\right)^2$$

$$d^2(\mathbf{r}_i, \mathbf{g}_r) = \sum_{j=1}^J \left(\frac{p_{ij}}{\sqrt{p_{\cdot j} p_{i \cdot}}} - \sqrt{p_{\cdot j}}\right)^2$$

Assim todas as coordenadas dos perfis de linha i e coluna j e seus respectivos centroides serão recalculados (ponderados) para serem projetados no novo espaço transformado (NAITO, 2007, p. 33), obtendo-se os perfis de linha \tilde{r} e coluna \tilde{j} e seus respectivos centroides $\tilde{\mathbf{g}}_r$ e $\tilde{\mathbf{g}}_c$.

Perfil de linha \tilde{r} :

$$\tilde{\mathbf{r}}_i^t = \left[\frac{p_{i1}}{\sqrt{p_{\cdot 1} p_{i \cdot}}} \quad \frac{p_{i2}}{\sqrt{p_{\cdot 2} p_{i \cdot}}} \quad \dots \quad \frac{p_{ij}}{\sqrt{p_{\cdot j} p_{i \cdot}}} \quad \dots \quad \frac{p_{ij}}{\sqrt{p_{\cdot j} p_{i \cdot}}} \right]$$

Perfil de coluna \tilde{j} :

$$\tilde{c}_j^t = \left[\frac{p_{1j}}{\sqrt{p_{1\cdot} \cdot p_{\cdot j}}} \quad \frac{p_{2j}}{\sqrt{p_{2\cdot} \cdot p_{\cdot j}}} \quad \dots \quad \frac{p_{ij}}{\sqrt{p_{i\cdot} \cdot p_{\cdot j}}} \quad \dots \quad \frac{p_{Ij}}{\sqrt{p_{I\cdot} \cdot p_{\cdot j}}} \right]$$

As novas matrizes de perfil de linha e coluna serão, respectivamente:

$$\tilde{R} = RD_c^{-\frac{1}{2}} \quad \text{e} \quad \tilde{C} = CD_r^{-\frac{1}{2}}$$

Cada componente dos novos centroides de linha e coluna serão respectivamente:

$$\tilde{g}_{rj} = \sum_{i=1}^I \frac{p_{ij}}{\sqrt{p_{\cdot j} p_{i\cdot}}} \cdot p_{i\cdot} = \frac{p_{\cdot j}}{\sqrt{p_{\cdot j}}} = \sqrt{p_{\cdot j}}$$

e

$$\tilde{g}_{ci} = \sum_{j=1}^J \frac{p_{ij}}{\sqrt{p_{i\cdot} p_{\cdot j}}} \cdot p_{\cdot j} = \frac{p_{i\cdot}}{\sqrt{p_{i\cdot}}} = \sqrt{p_{i\cdot}}$$

Assim, pode-se determinar matricialmente o centroide de linha e coluna do espaço transformado, como sendo respectivamente:

$$\tilde{g}_r = \tilde{R}^t \cdot m_r \quad \text{e} \quad \tilde{g}_c = \tilde{C}^t \cdot m_c$$

Recalculando os perfis e os centroides de linha e de coluna, referente ao exemplo ilustrativo da tabela 2, para o espaço transformado, tem-se:

Matriz de perfil de linha do espaço transformado:

$$\tilde{R} = RD_c^{-\frac{1}{2}} = \begin{bmatrix} \tilde{r}_1^t \\ \tilde{r}_2^t \\ \tilde{r}_3^t \\ \tilde{r}_4^t \end{bmatrix} = \begin{bmatrix} 0,548 & 0,555 & 0,641 \\ 0,329 & 1 & 0,385 \\ 0,822 & 0 & 0,962 \\ 1,494 & 0,151 & 0 \end{bmatrix},$$

Matriz de perfil de coluna do espaço transformado:

$$\tilde{C} = CD_r^{-\frac{1}{2}} = \begin{bmatrix} \tilde{c}_1^t \\ \tilde{c}_2^t \\ \tilde{c}_3^t \end{bmatrix} = \begin{bmatrix} 0,349 & 0,382 & 0,662 & 0,815 \\ 0,359 & 1,178 & 0 & 0,084 \\ 0,478 & 0,524 & 0,907 & 0 \end{bmatrix}$$

Centroides, no espaço transformado, das nuvens de linha N(I) e de coluna N(J) respectivamente:

$$\tilde{\mathbf{g}}_r = \tilde{\mathbf{R}}^t \cdot \mathbf{m}_r = \begin{bmatrix} \sqrt{0,37} \\ \sqrt{0,36} \\ \sqrt{0,27} \end{bmatrix} = \begin{bmatrix} 0,608 \\ 0,6 \\ 0,52 \end{bmatrix}$$

$$\tilde{\mathbf{g}}_c = \tilde{\mathbf{C}}^t \cdot \mathbf{m}_c = \begin{bmatrix} \sqrt{0,15} \\ \sqrt{0,50} \\ \sqrt{0,24} \\ \sqrt{0,11} \end{bmatrix} = \begin{bmatrix} 0,387 \\ 0,707 \\ 0,490 \\ 0,332 \end{bmatrix}$$

A Figura 21 ilustra a mudança do espaço original dos perfis de linha e seu centroide para o espaço transformado.

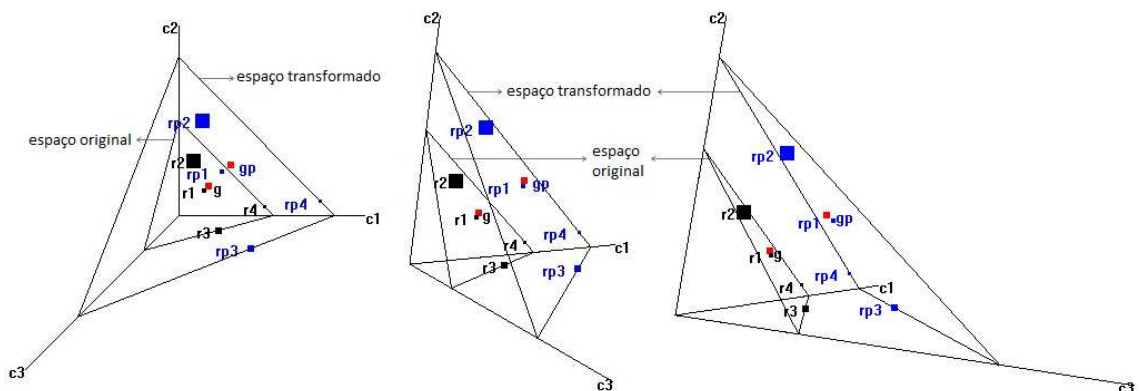


Figura 21 – Representação, sob 3 ângulos de visão, dos perfis de linha e o centroide no espaço original sob a métrica de qui-quadrado e no espaço transformado sob a métrica euclidiana

Pode-se perceber pela Figura 21 que o espaço transformado é estendido em relação ao espaço original.

Agora que os pontos (perfis) foram projetados em um espaço transformado no qual as distâncias entre os pontos deste espaço baseiam-se na tradicional distância euclidiana e não mais na distância de qui-quadrado, pode-se calcular a distância entre os pontos e o centroide de forma direta.

Assim pode-se facilmente identificar e medir quais pontos (perfis ou categorias) se distanciam do centroide, ponto este que retrata a independência entre as variáveis de linha e coluna da matriz original. Por exemplo, quanto mais o ponto que representa a categoria a1 da variável A (perfil de linha r1) se

distanciar do centroide e se aproximar de um dos eixos que representa a categoria b1 da variável B (eixo c1), conclui-se que mais evidente é a relação de dependência entre tais categorias. Além disso, é possível visualizar os pontos próximos um dos outros, o que demonstra similaridades entre eles.

Como o objetivo é medir a distância de cada perfil em relação ao seu centroide, pode-se transladar¹ a origem do sistema de eixos para o mesmo, necessitando então, recalculer as novas coordenadas dos pontos para o novo sistema de eixos centrado no centroide.

Ao se realizar a diferença das coordenadas de cada perfil pelas coordenadas do centroide, do espaço transformado, determinam-se as novas coordenadas centradas no centroide. A Figura 22 ilustra a translação da origem do sistema de eixos para o centroide, ocasionando a alteração das coordenadas dos pontos.

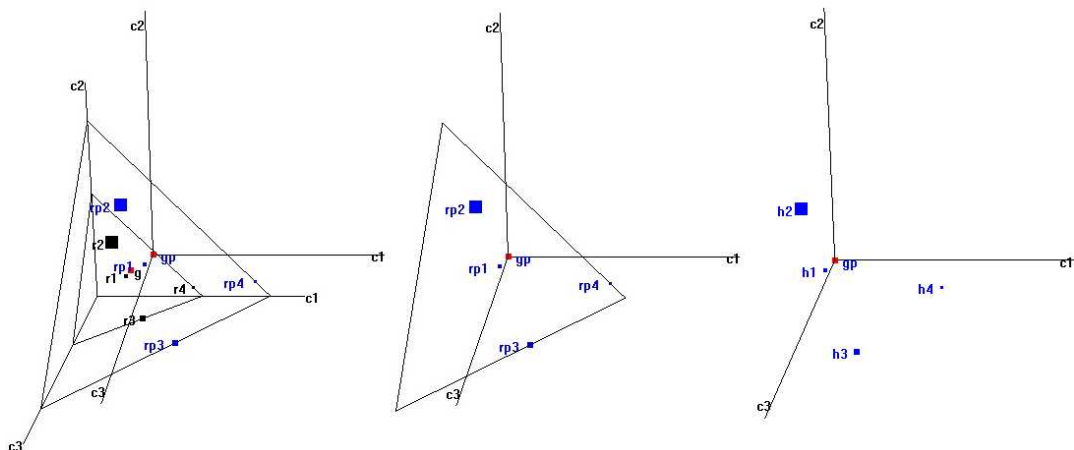


Figura 22: Translação da base para o centroide do espaço transformado

Estas novas coordenadas centralizadas no centroide serão fornecidas pela matriz H , que é a matriz de resíduos entre os perfis de linha e o centroide no espaço transformado:

$$H = \tilde{R} - \mathbf{1}_l \tilde{\mathbf{g}}_r^t$$

$$H = R D_c^{-\frac{1}{2}} - \mathbf{1}_l (\tilde{R}^t \cdot \mathbf{m}_r)^t$$

¹ Diz-se que um sistema de eixos foi transladado para um ponto P de um espaço n-dimensional quando os novos eixos são paralelos aos primitivos. Toda translação provoca uma mudança nas coordenadas de todos os pontos do espaço n-dimensional, tal que $P'=(P-Q)$, se os novos eixos estão no mesmo sentido que os eixos primitivos. (MURDOCH, 1975, p. 59)

$$\begin{aligned}
H &= RD_c^{-\frac{1}{2}} - \mathbf{1}_I \mathbf{m}_r^t RD_c^{-\frac{1}{2}} \\
H &= (D_r^{-1}P - \mathbf{1}_I \mathbf{m}_r^t D_r^{-1}P) D_c^{-\frac{1}{2}} \\
H &= D_r^{-1}(P - \mathbf{m}_r \mathbf{1}_I^t P) D_c^{-\frac{1}{2}} \\
H &= D_r^{-\frac{1}{2}} D_r^{-\frac{1}{2}} (P - \mathbf{m}_r \mathbf{m}_c^t) D_c^{-\frac{1}{2}} \\
H &= D_r^{-\frac{1}{2}} X
\end{aligned}$$

em que X será (NAITO, 2007, p. 43)

$$X = D_r^{-\frac{1}{2}} (P - \mathbf{m}_r \mathbf{m}_c^t) D_c^{-\frac{1}{2}}$$

pré-multiplicando X por $D_r^{\frac{1}{2}} D_r^{-\frac{1}{2}}$ a fórmula não se altera

$$\begin{aligned}
X &= D_r^{\frac{1}{2}} D_r^{-\frac{1}{2}} D_r^{-\frac{1}{2}} (P - \mathbf{m}_r \mathbf{m}_c^t) D_c^{-\frac{1}{2}} \\
X &= D_r^{\frac{1}{2}} (D_r^{-1}P - D_r^{-1} \mathbf{m}_r \mathbf{m}_c^t) D_c^{-\frac{1}{2}} \\
X &= D_r^{\frac{1}{2}} (R - \mathbf{1}_I \mathbf{m}_c^t) D_c^{-\frac{1}{2}}
\end{aligned}$$

voltando em H , tem-se que

$$\begin{aligned}
H &= D_r^{-\frac{1}{2}} X \\
H &= D_r^{-\frac{1}{2}} D_r^{\frac{1}{2}} (R - \mathbf{1}_I \mathbf{m}_c^t) D_c^{-\frac{1}{2}} \\
H &= (R - \mathbf{1}_I \mathbf{m}_c^t) D_c^{-\frac{1}{2}}
\end{aligned}$$

Logo,

$$X = D_r^{\frac{1}{2}} H$$

Portanto a matriz X contém os perfis de linha no espaço transformado, ponderado pela raiz de massa.

Pode-se perceber que a matriz H é formada em suas linhas pelas coordenadas dos perfis de linha, já no espaço transformado, centradas no centroide. Para se calcular a inércia de um perfil de linha qualquer, deve-se calcular a distância quadrática que possui em relação ao centroide, usando a métrica da distância de qui-quadrado e ponderá-la pela massa do perfil, $In(r_i) = m_i \cdot d(\tilde{r}_i, \tilde{g}_r)^2$. Ao se somar a inércia de todos os perfis, tem-se a inércia da nuvem de linha $N(I)$, $In(I) = \sum_{i=1}^I m_i \cdot d(\tilde{r}_i, \tilde{g}_r)^2$.

A matriz H gera de forma direta as distâncias $d(\tilde{r}_i, \tilde{g}_r)$ entre os perfis de linha e o centroide do espaço transformado, já usando a métrica de qui-quadrado, e, ao se ponderar suas coordenadas pela raiz da massa do perfil, encontram-se novos pontos " x_i " no espaço transformado que facilitam tremendamente o cálculo final da inércia, visto que a sua distância à origem (centroide) será $d(x_i, \tilde{g}_r) = \sqrt{m_i} \cdot d(\tilde{r}_i, \tilde{g}_r) = \sqrt{m_i \cdot d^2(\tilde{r}_i, \tilde{g}_r)} = \sqrt{In(r_i)}$.

A grande vantagem desses novos pontos é que o cálculo da inércia se dará de forma direta, bastando fazer o quadrado da distância do ponto x_i ao centroide. As coordenadas desses novos pontos x_i serão fornecidas pela matriz X , já apresentada.

A seguir calcula-se a matriz H e a matriz X para os dados do exemplo ilustrativo da tabela 2 que posteriormente serão projetados graficamente.

$$H = \tilde{R} - \mathbf{1}_I \tilde{g}_r^t$$

$$H = \begin{bmatrix} 0,548 & 0,555 & 0,641 \\ 0,329 & 1 & 0,385 \\ 0,822 & 0 & 0,962 \\ 1,494 & 0,151 & 0 \end{bmatrix} - \begin{bmatrix} 0,608 & 0,6 & 0,52 \\ 0,608 & 0,6 & 0,52 \\ 0,608 & 0,6 & 0,52 \\ 0,608 & 0,6 & 0,52 \end{bmatrix}$$

$$H = \begin{bmatrix} h_1^t \\ h_2^t \\ h_3^t \\ h_4^t \end{bmatrix} = \begin{bmatrix} -0,06 & -0,044 & 0,122 \\ -0,279 & 0,4 & -0,135 \\ 0,214 & -0,6 & 0,443 \\ 0,886 & -0,448 & -0,52 \end{bmatrix}$$

$$X = \begin{bmatrix} \sqrt{0,15} & 0 & 0 & 0 \\ 0 & \sqrt{0,50} & 0 & 0 \\ 0 & 0 & \sqrt{0,24} & 0 \\ 0 & 0 & 0 & \sqrt{0,11} \end{bmatrix} \begin{bmatrix} -0,06 & -0,044 & 0,122 \\ -0,279 & 0,4 & -0,135 \\ 0,214 & -0,6 & 0,443 \\ 0,886 & -0,448 & -0,52 \end{bmatrix}$$

$$X = \begin{bmatrix} x_1^t \\ x_2^t \\ x_3^t \\ x_4^t \end{bmatrix} = \begin{bmatrix} -0,023 & -0,017 & 0,047 \\ -0,198 & 0,283 & -0,095 \\ 0,105 & -0,294 & 0,217 \\ 0,294 & -0,149 & -0,172 \end{bmatrix}$$

A Figura 23 ilustra a projeção dos perfis h_i e x_i , sendo este último ponderado pela raiz de sua massa.

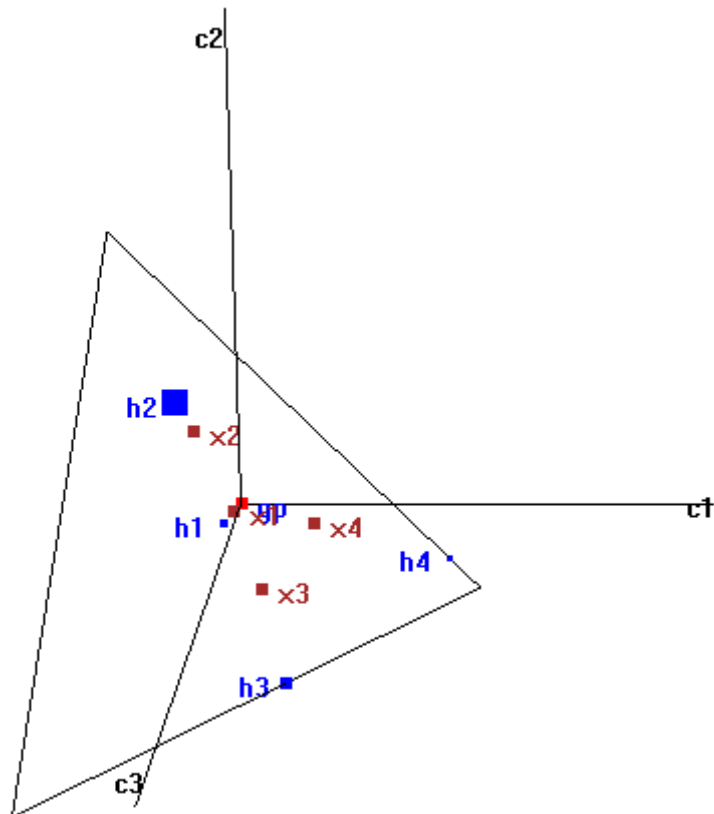


Figura 23: Representação comparativa dos perfis h_i e x_i no espaço transformado

Pode-se perceber pela Figura 23 que o deslocamento do perfil h_4 para próximo do centroide foi mais acentuado do que o perfil h_2 , visto que este último (massa=0,50) possui uma massa maior que o perfil h_4 (massa=0,11).

Agora as distâncias dos novos pontos podem ser medidas naturalmente pela métrica euclidiana, visto que a mesma foi compensada por sua massa, ou seja, pela sua representatividade ou relevância, e também já foi equilibrada por sua relatividade (distância de qui-quadrado).

Sendo assim, ao se realizar o produto interno entre os vetores coluna da matriz X , obtém-se a matriz S que pode ser considerada uma matriz de

variâncias e covariâncias (PAMPLONA, 1998, p. 20) dos resíduos dos perfis de linha ponderados no espaço transformado.

$$X^t = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_J \end{bmatrix} \text{ e } X = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_J \end{bmatrix}.$$

$$X^t X = S = \begin{bmatrix} \|\mathbf{c}_1\|^2 & \langle \mathbf{c}_1, \mathbf{c}_2 \rangle & \dots & \langle \mathbf{c}_1, \mathbf{c}_J \rangle \\ \langle \mathbf{c}_2, \mathbf{c}_1 \rangle & \|\mathbf{c}_2\|^2 & \dots & \langle \mathbf{c}_2, \mathbf{c}_J \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{c}_J, \mathbf{c}_1 \rangle & \langle \mathbf{c}_J, \mathbf{c}_2 \rangle & \dots & \|\mathbf{c}_J\|^2 \end{bmatrix}$$

Sendo assim, pode-se escrever cada elemento da matriz S como sendo,

$$\langle \mathbf{c}_j, \mathbf{c}_{j'} \rangle = \sum_{i=1}^I (k_{ij} - k_{\tilde{g}_{rj}}) (k_{ij'} - k_{\tilde{g}_{rj'}}) = \sum_{i=1}^I k_{ij} k_{ij'}$$

em que,

k_{ij} é a coordenada do ponto x_i (linha i da matriz X) sobre o j -ésimo eixo centrado no centroide.

$k_{\tilde{g}_{rj}} = 0$, pois esta é a coordenada do centroide \tilde{g}_r sobre o j -ésimo eixo centrado no centroide.

Pode-se observar a semelhança da fórmula anterior com a covariância, e quando j for igual a j' tem-se a variância, justificando o termo de matriz de variâncias e covariâncias atribuída à matriz S .

Como a distância de cada ponto x_i ao centroide $d(x_i, \tilde{g}_r)$ é igual à raiz da inércia do perfil de linha i , $\sqrt{In(r_i)}$, então pode-se projetar estas distâncias nos eixos de base.

$$\sqrt{In(r_i)} = \left\| \sum_{j=1}^J (Proj x_i)_j \right\| = \left\| \sum_{j=1}^J \vec{k}_{ij} \right\|$$

sendo $(Proj x_i)_j$ a projeção do ponto x_i sobre o j -ésimo eixo e \vec{k}_{ij} o vetor centrado na origem do espaço transformado com extremidade em k_{ij} .

Fazendo $p_j x_{ij} = (Proj x_i)_j$, pode-se ter outra visão da matriz S , como:

$$S = \begin{bmatrix} \sum_{i=1}^I \|p_j x_{i1}\|^2 & \sum_{i=1}^I \|p_j x_{i1}\| \|p_j x_{i2}\| & \cdots & \sum_{i=1}^I \|p_j x_{i1}\| \|p_j x_{iJ}\| \\ \sum_{i=1}^I \|p_j x_{i2}\| \|p_j x_{i1}\| & \sum_{i=1}^I \|p_j x_{i2}\|^2 & \cdots & \sum_{i=1}^I \|p_j x_{i2}\| \|p_j x_{iJ}\| \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^I \|p_j x_{iJ}\| \|p_j x_{i1}\| & \sum_{i=1}^I \|p_j x_{iJ}\| \|p_j x_{i2}\| & \cdots & \sum_{i=1}^I \|p_j x_{iJ}\|^2 \end{bmatrix}$$

$$S = \begin{bmatrix} In(j=1) & \sqrt{In(j=1)}\sqrt{In(j=2)} & \cdots & \sqrt{In(j=1)}\sqrt{In(j=J)} \\ \sqrt{In(j=2)}\sqrt{In(j=1)} & In(j=2) & \cdots & \sqrt{In(j=2)}\sqrt{In(j=J)} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{In(j=J)}\sqrt{In(j=1)} & \sqrt{In(j=J)}\sqrt{In(j=2)} & \cdots & In(j=J) \end{bmatrix}$$

Para o exemplo da tabela 2, a matriz S é dada por:

$$X^t X = S = \begin{bmatrix} -0,023 & -0,198 & 0,105 & 0,294 \\ -0,017 & 0,283 & -0,294 & -0,149 \\ 0,047 & -0,095 & 0,217 & -0,172 \end{bmatrix} \cdot \begin{bmatrix} -0,023 & -0,017 & 0,047 \\ -0,198 & 0,283 & -0,095 \\ 0,105 & -0,294 & 0,217 \\ 0,294 & -0,149 & -0,172 \end{bmatrix}$$

$$S = \begin{bmatrix} 0,137 & -0,13 & -0,01 \\ -0,13 & 0,189 & -0,066 \\ -0,01 & -0,066 & 0,088 \end{bmatrix}$$

O traço de S fornece a inércia da nuvem $N(I)$, sendo os elementos de sua diagonal a inércia projetada sobre os eixos que formam sua base no espaço transformado. Como cada eixo de base representa uma categoria de coluna da matriz de dados W , então os elementos da diagonal da matriz S também serão a inércia dos perfis de coluna.

Determinando as matrizes K , Y e Z , que correspondem respectivamente às matrizes H , X e S , associadas à nuvem de coluna $N(J)$.

$$K = \tilde{C} - \mathbf{1}_J \tilde{\mathbf{g}}_c^t$$

$$K = D_c^{-\frac{1}{2}} D_c^{-\frac{1}{2}} (P^t - \mathbf{m}_c \mathbf{m}_r^t) D_r^{-\frac{1}{2}}$$

$$K = D_c^{-\frac{1}{2}} Y$$

em que Y será

$$Y = D_c^{-\frac{1}{2}} (P^t - \mathbf{m}_c \mathbf{m}_r^t) D_r^{-\frac{1}{2}} \mathbf{e}$$

$$Y = D_c^{-\frac{1}{2}} (C - \mathbf{1}_J \mathbf{m}_r^t) D_r^{-\frac{1}{2}}$$

Voltando em K , tem-se que

$$K = D_c^{-\frac{1}{2}} Y$$

$$K = (C - \mathbf{1}_J \mathbf{m}_r^t) D_r^{-\frac{1}{2}}$$

Logo

$$Y = D_c^{-\frac{1}{2}} K$$

Comparando as matrizes X e Y percebe-se que

$$X = D_r^{-\frac{1}{2}} (P - \mathbf{m}_r \mathbf{m}_c^t) D_c^{-\frac{1}{2}}$$

$$Y = X^t = D_c^{-\frac{1}{2}} (P^t - \mathbf{m}_c \mathbf{m}_r^t) D_r^{-\frac{1}{2}}$$

Assim a matriz $Z = Y^t Y = X X^t$ (PAMPLONA, 1998, p. 36).

Determinando as matrizes K , Y e Z para o exemplo da tabela 2, tem-se:

$$K = \tilde{C} - \mathbf{1}_J \tilde{\mathbf{g}}_c^t$$

$$K = \begin{bmatrix} 0,349 & 0,382 & 0,662 & 0,815 \\ 0,359 & 1,178 & 0 & 0,084 \\ 0,478 & 0,524 & 0,907 & 0 \end{bmatrix} - \begin{bmatrix} 0,387 & 0,707 & 0,49 & 0,332 \\ 0,387 & 0,707 & 0,49 & 0,332 \\ 0,387 & 0,707 & 0,49 & 0,332 \end{bmatrix}$$

$$K = \begin{bmatrix} -0,038 & -0,325 & 0,172 & 0,483 \\ -0,029 & 0,471 & -0,49 & -0,248 \\ 0,091 & -0,183 & 0,417 & -0,331 \end{bmatrix}$$

$$Y = X^t = \begin{bmatrix} -0,023 & -0,198 & 0,105 & 0,294 \\ -0,017 & 0,283 & -0,294 & -0,149 \\ 0,047 & -0,095 & 0,217 & -0,172 \end{bmatrix}$$

$$XX^t = Z = \begin{bmatrix} -0,023 & -0,017 & 0,047 \\ -0,198 & 0,283 & -0,095 \\ 0,105 & -0,294 & 0,217 \\ 0,294 & -0,149 & -0,172 \end{bmatrix} \cdot \begin{bmatrix} -0,023 & -0,198 & 0,105 & 0,294 \\ -0,017 & 0,283 & -0,294 & -0,149 \\ 0,047 & -0,095 & 0,217 & -0,172 \end{bmatrix}$$

$$Z = \begin{bmatrix} 0,003 & -0,005 & 0,013 & -0,012 \\ -0,005 & 0,128 & -0,124 & -0,083 \\ 0,013 & -0,124 & 0,144 & 0,037 \\ -0,012 & -0,08 & 0,371 & 0,138 \end{bmatrix}$$

Pode-se perceber que a diagonal da matriz S traz a inércia da nuvem de linha $N(I)$ sobre os eixos. Tais inércias coincidem com a inércia dos perfis de coluna. Assim como a diagonal da matriz Z fornece a inércia da nuvem de coluna $N(J)$ sobre os eixos que coincidem com a inércia dos perfis de linha. O traço de ambas as matrizes são iguais e fornecem a inércia total das nuvens.

Não serão projetadas as matrizes K e Y , pois a nuvem de coluna está no \mathbb{R}^4 e portanto não pode ser visualizada.

2.2.11 Redução de dimensionalidade

O objetivo final da análise de correspondência é reduzir a dimensionalidade da nuvem de dados com o mínimo possível de perda de informação, ou seja, retendo o máximo possível da inércia (variabilidade). Normalmente consideram-se dois ou três eixos (subespaço ótimo bi ou tridimensional) favorecendo assim uma interpretação facilitada baseada apenas na inspeção gráfica dos relacionamentos entre os pontos (categorias) a partir das suas distâncias.

Assim, para a determinação do subespaço ótimo é preciso determinar uma base ortonormal do espaço transformado cujos vetores minimizem o erro provocado ao se usar a projeção dos pontos da nuvem sobre si, isto é, que maximizem as projeções dos pontos com a finalidade de reter o máximo possível da inércia. Para tanto é preciso determinar as projeções ortogonais dos pontos da nuvem sobre os vetores da base (item 2.1.5.1).

Logo, o que se deseja determinar é um vetor unitário v sobre o qual serão feitas as projeções dos perfis de linha x_i da matriz X . Como v é unitário então pode-se determinar a norma da projeção a partir do produto interno das linhas de X com v , isto é Xv . Fazendo $(Xv)^t Xv$, este produto interno representa o somatório das normas quadráticas das projeções, que nada mais é que a inércia no eixo definido por v , pois cada elemento de X representa a raiz da inércia.

Como $(Xv)^t Xv = v^t X^t Xv = v^t S v$ é uma forma quadrática, pode-se aplicar o teorema dos eixos principais (item 2.1.6) com objetivo de definir uma nova base ortonormal $\beta = \{v_1, v_2, \dots, v_j\}$ centrada no centroide do espaço transformado que, a partir da qual permitirá definir um subespaço ótimo (item 2.1.7).

Para determinar β deve-se proceder à decomposição espectral sobre a matriz simétrica (variâncias e covariâncias) S conforme mostrado a seguir:

$$S = V D_\lambda V^t$$

D_λ é a matriz diagonal contendo os autovalores de S em ordem decrescente $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_j = 0$;

V é matriz de autovetores associados aos autovalores λ de tal forma que λ_j está associado a v_j .

$$\lambda_1 \Rightarrow v_1$$

$$\lambda_2 \Rightarrow v_2$$

$$\vdots$$

$$\lambda_j \Rightarrow v_j$$

Portanto, λ_j representa a inércia captada por cada eixo v_j e será útil para a interpretação da proporção da inércia que é explicada pelos eixos.

Para o exemplo ilustrativo da tabela 2, a decomposição espectral das matrizes S e Z , tem-se:

$$S = X^t X = V D_\lambda V^t$$

$$S = \begin{bmatrix} 0,590 & 0,53 & 0,608 \\ -0,779 & 0,18 & 0,6 \\ 0,209 & -0,829 & 0,52 \end{bmatrix} \begin{bmatrix} 0,305 & 0 & 0 \\ 0 & 0,109 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0,59 & -0,779 & 0,209 \\ 0,53 & 0,18 & -0,829 \\ 0,608 & 0,6 & 0,52 \end{bmatrix}$$

A Figura 24 ilustra a representação dos vetores v_1, v_2, v_3 no espaço transformado.

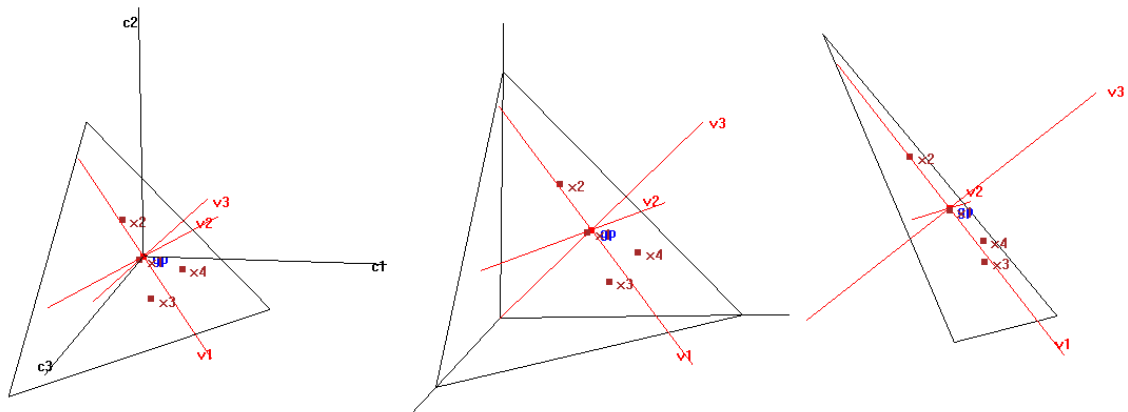


Figura 24: Representação eixos principais v_1, v_2, v_3 no espaço transformado

Observa-se que v_3 é o próprio centroide e está associado ao autovalor zero e, portanto, não capta nenhuma variabilidade da nuvem $N(I)$. Isto se justifica, pois v_3 é perpendicular ao subespaço formado por v_1 e v_2 . Como os autovetores associados a autovalores zero não captam nenhuma informação da nuvem de pontos, os mesmos podem ser descartados, o que resulta na decomposição espectral na forma reduzida, conforme apresentado a seguir:

$$S = \begin{bmatrix} 0,590 & 0,53 \\ -0,779 & 0,18 \\ 0,209 & -0,829 \end{bmatrix} \begin{bmatrix} 0,305 & 0 \\ 0 & 0,109 \end{bmatrix} \begin{bmatrix} 0,59 & -0,779 & 0,209 \\ 0,53 & 0,18 & -0,829 \end{bmatrix}$$

Fazendo-se a decomposição espectral já na forma reduzida sobre Z , tem-se

$$Z = XX^t = UD_\lambda U^T$$

$$Z = \begin{bmatrix} -0,017 & 0,165 \\ 0,647 & -0,076 \\ -0,609 & 0,536 \\ -0,459 & -0,824 \end{bmatrix} \begin{bmatrix} 0,305 & 0 \\ 0 & 0,109 \end{bmatrix} \begin{bmatrix} -0,017 & 0,647 & -0,609 & -0,459 \\ 0,165 & -0,076 & 0,536 & -0,824 \end{bmatrix}$$

Neste exemplo, o subespaço ótimo de dimensão dois para as linhas e colunas de X é o subespaço gerado pelos dois vetores de V e de U , respectivamente, associados aos dois maiores autovalores.

2.2.12 Relação entre linhas e colunas

Até esse ponto foi apresentado como estabelecer um subespaço ótimo, tanto para nuvem de linha quanto para nuvem de coluna, utilizando os eixos principais que capturaram a maior parte da inércia da nuvem em questão. Nesta subseção, serão apresentados os principais aspectos que relacionam as nuvens de linha e coluna.

Os eixos principais da nuvem de linha e coluna foram encontrados através da decomposição espectral da matriz X^tX e XX^t respectivamente obtendo-se assim as matrizes V e U (GREENACRE; BLASIUS, 2006, p. 13):

$$X^tX = VD_\lambda V^t$$

$$XX^t = UD_\lambda U^t$$

em que V e U são matrizes contendo os autovetores que formaram os eixos principais da nuvem linha e coluna respectivamente. Assim pode-se estabelecer que

$$X^tXV = VD_\lambda \tag{4}$$

e

$$XX^tU = UD_\lambda \tag{5}$$

Pré-multiplicando por X em (4), tem-se:

$$XX^tXV = XVD_\lambda \tag{6}$$

Comparando (6) com (5) percebe-se que XV é uma matriz de autovetores de XX^t , portanto é proporcional a U . Sendo assim, pode-se estabelecer que:

$$XV = UD_a \tag{7}$$

portanto,

$$X = UD_aV^t \tag{8}$$

Assim, deve-se determinar D_a em relação à D_λ que é conhecido.

$$X^t = VD_a U^t \quad (9)$$

substituindo (9) em (5), tem-se

$$XVD_a U^t U = UD_\lambda \quad (10)$$

substituindo (7) em (10), e pré-multiplicando por U^t , tem-se

$$U^t U D_a D_a U^t U = U^t U D_\lambda$$

em que,

$$D_a^2 = D_\lambda$$

portanto,

$$D_a = D_\lambda^{\frac{1}{2}} \quad (11)$$

Substituindo (11) em (8), tem-se a decomposição em valores singulares (DVS) da matriz X

$$X = UD_\lambda^{\frac{1}{2}} V^t \quad (12)$$

Daqui em diante $D_\lambda^{\frac{1}{2}} = D_\sigma$, então,

$$X = UD_\sigma V^t \quad (13)$$

ou

$$Y = X^t = VD_\sigma U^t \quad (14)$$

Portanto, realizando a DVS sobre a matriz X obtém-se diretamente os eixos principais da nuvem de linha N(I) por meio da matriz V , os eixos principais da nuvem de coluna N(J) por meio da matriz U , como também a correspondente inércia captada nesses eixos principais, fornecido pelo quadrado dos valores singulares (diagonal) de D_σ .

É importante salientar que ao se calcular U e V de forma independente e utilizá-los para compor a matriz X , podem ocorrer diferenças nos sinais entre as linhas da matriz X e daquela obtida por $UD_\sigma V^t$. Isto porque há dois autovetores unitários associados ao mesmo autovalor, com a mesma direção, porém com sentido opostos. Pode-se corrigir tal distorção calculando-se U em relação à V ou vice-versa, conforme apresentado a seguir:

Pós-multiplicando por V , seguido de D_σ^{-1} em (13), tem-se

$$U = XVD_\sigma^{-1}$$

ou pós-multiplicando por U , seguido de D_σ^{-1} em (14), tem-se

$$V = X^tUD_\sigma^{-1}$$

Recalculando U em relação à V , referente ao exemplo da tabela 2, tem-se,

$$U = XVD_\sigma^{-1}$$

$$U = \begin{bmatrix} -0,023 & -0,017 & 0,047 \\ -0,198 & 0,283 & -0,095 \\ 0,105 & -0,294 & 0,217 \\ 0,294 & -0,149 & -0,172 \end{bmatrix} \begin{bmatrix} 0,590 & 0,53 \\ -0,779 & 0,18 \\ 0,209 & -0,829 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0,552 & 1 \\ 0 & 0,330 \end{bmatrix}$$

$$U = \begin{bmatrix} 0,017 & -0,165 \\ -0,647 & 0,076 \\ 0,609 & -0,536 \\ 0,459 & 0,824 \end{bmatrix}$$

Assim, a decomposição em valores singulares de matriz X será

$$X = UD_\sigma V^t$$

$$X = \begin{bmatrix} 0,017 & -0,165 \\ -0,647 & 0,076 \\ 0,609 & -0,536 \\ 0,459 & 0,824 \end{bmatrix} \begin{bmatrix} 0,552 & 0 \\ 0 & 0,330 \end{bmatrix} \begin{bmatrix} 0,59 & -0,779 & 0,209 \\ 0,53 & 0,18 & -0,829 \end{bmatrix}$$

2.2.13 As coordenadas nos eixos principais

Conforme já apresentado no subitem 2.1.2, pode-se definir as coordenadas de um vetor sobre uma base ortonormal através do produto interno entre o vetor que se deseja projetar e vetores da base. Assim, para se determinar as coordenadas dos perfis de linha da matriz X (ponderados pela massa, no espaço transformado) sobre a nova base $\beta = \{v_1, v_2, \dots, v_J\}$ (obtida através da matriz ortonormal V) bastaria realizar o produto interno entre as linhas da matriz X e as colunas da matriz V .

$$(\text{proj } X)_\beta = (XV)V$$

$$\tilde{F} = XV$$

Em AC procura-se representar os resíduos dos perfis de linha e de coluna em um subespaço ótimo, porém, esse subespaço é influenciado pela métrica de qui-quadrado e pela ponderação da massa, o que inviabiliza a redução direta de dimensionalidade sob tais características. Assim, até esse ponto, transformou-se os resíduos dos perfis de linha originais (matriz $R - \mathbf{1}_I \mathbf{m}_c^t$) na matriz X para que fosse possível aplicar técnicas tais como decomposição espectral, projeções ortogonais entre outras que são baseadas na métrica euclidiana.

Como X pode ser escrito como sendo o resíduo entre os perfis de linha e o centroide, ponderado por sua massa $D_r^{-\frac{1}{2}}$ e equilibrado pela distancia de qui-quadrado $D_c^{-\frac{1}{2}}$, pode-se fazer uma transformação inversa para determinar os vetores do subespaço ótimo dos resíduos dos perfis de linha. A mesma analogia pode ser feita para os resíduos dos perfis de coluna, conforme mostrado abaixo.

$$X = D_r^{-\frac{1}{2}}(R - \mathbf{1}_I \mathbf{m}_c^t)D_c^{-\frac{1}{2}}$$

$$Y = X^t = D_c^{-\frac{1}{2}}(C - \mathbf{1}_J \mathbf{m}_r^t)D_r^{-\frac{1}{2}}$$

$$X = UD_\sigma V^t$$

$$X^t = VD_\sigma U^t$$

$$(R - \mathbf{1}_I \mathbf{m}_c^t) = D_r^{-\frac{1}{2}}UD_\sigma V^t D_c^{-\frac{1}{2}}$$

$$(C - \mathbf{1}_J \mathbf{m}_r^t) = D_c^{-\frac{1}{2}}VD_\sigma U^t D_r^{-\frac{1}{2}}$$

$$N = D_r^{-\frac{1}{2}}U$$

$$E = D_c^{-\frac{1}{2}}V$$

$$M = D_c^{-\frac{1}{2}}V$$

$$L = D_r^{-\frac{1}{2}}U$$

$$(R - \mathbf{1}_I \mathbf{m}_c^t) = ND_\sigma M^t$$

$$(C - \mathbf{1}_J \mathbf{m}_r^t) = ED_\sigma L^t$$

$$U^t U = N^t D_r^{-\frac{1}{2}} D_r^{-\frac{1}{2}} N = N^t D_r N = I$$

$$U^t U = L^t D_r^{-\frac{1}{2}} D_r^{-\frac{1}{2}} L = L^t D_r^{-1} L = I$$

$$V^t V = M^t D_c^{-\frac{1}{2}} D_c^{-\frac{1}{2}} M = M^t D_c^{-1} M = I$$

$$V^t V = E^t D_c^{-\frac{1}{2}} D_c^{-\frac{1}{2}} E = E^t D_c E = I$$

Como mostrado acima, tanto $(R - \mathbf{1}_I \mathbf{m}_c^t)$ como $(C - \mathbf{1}_J \mathbf{m}_r^t)$ foram decompostos em valores singulares generalizados pelas métricas de qui-quadrado (D_c^{-1} e D_r^{-1} respectivamente) e pela massa (D_r e D_c), que era o almejado. Desta forma, as coordenadas das projeções dos resíduos dos perfis de linha em M são dadas por $F = ND_\sigma$ e as coordenadas dos resíduos dos perfis de coluna em L são dadas por $G = ED_\sigma$. A seguir demonstra-se a relação para os resíduos dos perfis de linha.

Sendo

$$M^t D_c^{-1} M = M^t D_c^{-\frac{1}{2}} D_c^{-\frac{1}{2}} M = I,$$

e fazendo $O = D_c^{-\frac{1}{2}} M$, então $O^t O = I$, o que implica que O é ortornormal.

Colocando-se o resíduo do perfil de linha sob a métrica euclidiana tem-se $(R - \mathbf{1}_I \mathbf{m}_c^t) D_c^{-\frac{1}{2}}$. As coordenadas deste resíduo sobre O podem ser obtidas por produto interno, conforme já mencionado no item 2.1.2:

$$F = (R - \mathbf{1}_I \mathbf{m}_c^t) D_c^{-\frac{1}{2}} \cdot O$$

$$F = (R - \mathbf{1}_I \mathbf{m}_c^t) D_c^{-\frac{1}{2}} \cdot D_c^{-\frac{1}{2}} M$$

$$F = (R - \mathbf{1}_I \mathbf{m}_c^t) D_c^{-1} M$$

como

$$(R - \mathbf{1}_I \mathbf{m}_c^t) = ND_\sigma M^t$$

e pós-multiplicando por $D_c^{-1} M$, obtém-se que

$$(R - \mathbf{1}_I \mathbf{m}_c^t) D_c^{-1} M = ND_\sigma M^t D_c^{-1} M = ND_\sigma.$$

Logo,

$$F = ND_\sigma$$

Há inúmeras maneiras de apresentação das matrizes de coordenadas

F e G

$$\begin{array}{ll} F = ND_{\sigma} & G = ED_{\sigma} \\ F = D_r^{-\frac{1}{2}}UD_{\sigma} & G = D_c^{-\frac{1}{2}}VD_{\sigma} \\ F = D_r^{-\frac{1}{2}}XV & G = D_c^{-\frac{1}{2}}X^tU \\ F = D_r^{-1}LD_{\sigma} & G = D_c^{-1}MD_{\sigma} \end{array}$$

X também pode ser escrito como sendo o resíduo entre as frequências relativas observadas (matriz de correspondência P) e esperadas (matriz $\mathbf{m}_r\mathbf{m}_c^t$ de frequências relativas esperadas para uma situação de independência entre linhas e colunas da matriz). Pode-se fazer a mesma analogia para matriz Y .

$$X = D_r^{-\frac{1}{2}}(P - \mathbf{m}_r\mathbf{m}_c^t)D_c^{-\frac{1}{2}} \qquad Y = X^t = D_c^{-\frac{1}{2}}(P^t - \mathbf{m}_c\mathbf{m}_r^t)D_r^{-\frac{1}{2}}$$

$$\begin{array}{ll} X = UD_{\sigma}V^t & X^t = VD_{\sigma}U^t \\ (P - \mathbf{m}_r\mathbf{m}_c^t) = D_r^{\frac{1}{2}}UD_{\sigma}V^tD_c^{\frac{1}{2}} & (P^t - \mathbf{m}_c\mathbf{m}_r^t) = D_c^{\frac{1}{2}}VD_{\sigma}U^tD_r^{\frac{1}{2}} \\ A = D_r^{\frac{1}{2}}U & B = D_c^{\frac{1}{2}}V \\ B = D_c^{\frac{1}{2}}V & A = D_r^{\frac{1}{2}}U \\ (P - \mathbf{m}_r\mathbf{m}_c^t) = AD_{\sigma}B^t & (P^t - \mathbf{m}_c\mathbf{m}_r^t) = BD_{\sigma}A^t \end{array}$$

$$\begin{array}{ll} U^tU = A^tD_r^{-\frac{1}{2}}D_r^{-\frac{1}{2}}A = A^tD_r^{-1}A = I & U^tU = A^tD_r^{-\frac{1}{2}}D_r^{-\frac{1}{2}}A = A^tD_r^{-1}A = I \\ V^tV = B^tD_c^{-\frac{1}{2}}D_c^{-\frac{1}{2}}B = B^tD_c^{-1}B = I & V^tV = B^tD_c^{-\frac{1}{2}}D_c^{-\frac{1}{2}}B = B^tD_c^{-1}B = I \end{array}$$

A grande vantagem desta abordagem é a facilidade de cálculos, pois uma vez calculadas as matrizes envolvidas na nuvem de linha, tem-se automaticamente calculadas as matrizes envolvidas na nuvem de coluna.

A partir desta abordagem tem-se que

$$F = D_r^{-1}AD_\sigma$$

$$G = D_c^{-1}BD_\sigma$$

visto que $L = A$ e $M = B$.

As fórmulas apresentadas acima para F e G são chamadas de coordenadas principais da nuvem de linha e da nuvem de coluna respectivamente.

Ao se calcular F e G de forma independente pode levar a “erros” de sinais, como assegura Greenacre (1984, p. 93). Essas coordenadas serão utilizadas para plotar gráficos nos quais as proximidades entre os pontos são significativas e uma inversão nos sinais pode levar a marcação do ponto em um quadrante não compatível, o que prejudicaria a interpretação dos resultados. Por isto, são estabelecidas fórmulas de transição, as quais permitem calcular G a partir de F e vice-versa. Tais fórmulas não são necessárias se U foi calculado a partir de V ou vice-versa.

Um resultado importante para a obtenção das fórmulas de transição é o que foi observado ao fazer o desenvolvimento do exemplo no item 2.2.11: a matriz XX^t (ou X^tX) tem o centroide de coluna (ou linha) como autovetor associado ao autovalor 0. Este fato é demonstrado a seguir:

$$\begin{aligned} XX^t \left(D_r^{-\frac{1}{2}} \mathbf{m}_r \right) &= X \left[D_c^{\frac{1}{2}} (C - \mathbf{1}_J \mathbf{m}_r^t) D_r^{-\frac{1}{2}} \right] \left(D_r^{\frac{1}{2}} \mathbf{1}_I \right) \\ &= X D_c^{\frac{1}{2}} (C - \mathbf{1}_J \mathbf{m}_r^t) \mathbf{1}_I \\ &= X D_c^{\frac{1}{2}} (C \mathbf{1}_I - \mathbf{1}_J \mathbf{m}_r^t \mathbf{1}_I) \\ &= X D_c^{\frac{1}{2}} (\mathbf{1}_J - \mathbf{1}_J) \\ &= X D_c^{\frac{1}{2}} \mathbf{0}_J \\ &= \mathbf{0}_I = 0 \left(D_r^{-\frac{1}{2}} \mathbf{m}_r \right) \end{aligned}$$

Isto quer dizer que o centroide de coluna do espaço transformado é um dos autovetores da decomposição espectral não reduzida de XX^t .

Semelhantemente, demonstra-se que o centroide de linha é autovetor da matriz X^tX associado ao autovalor 0.

Pode-se agora passar à obtenção das fórmulas de transição.

Sendo $(P^t - \mathbf{m}_c \mathbf{m}_r^t) D_r^{-1} A = B D_\sigma$ e $D_r^{-1} A = F D_\sigma^{-1}$, então

$$G = D_c^{-1} B D_\sigma$$

$$G = D_c^{-1} (P^t - \mathbf{m}_c \mathbf{m}_r^t) D_r^{-1} A$$

$$G = D_c^{-1} P^t D_r^{-1} A - D_c^{-1} \mathbf{m}_c \mathbf{m}_r^t D_r^{-1} A$$

$$G = (D_c^{-1} P^t) (D_r^{-1} A) - (D_c^{-1} \mathbf{m}_c) \mathbf{m}_r^t (D_r^{-1} A)$$

$$G = C F D_\sigma^{-1} - \mathbf{1}_J \mathbf{m}_r^t D_r^{-1} A$$

$$G = C F D_\sigma^{-1} - \mathbf{1}_J \mathbf{m}_r^t D_r^{-1} D_r^{\frac{1}{2}} U$$

$$G = C F D_\sigma^{-1} - \left(\mathbf{1}_J \mathbf{m}_r^t D_r^{-\frac{1}{2}} \right) U$$

Mas $\left(\mathbf{1}_J \mathbf{m}_r^t D_r^{-\frac{1}{2}} \right)$ é uma matriz em que cada linha é o centroide de coluna do espaço transformado, que como já visto, é um dos autovetores da decomposição espectral não reduzida de XX^t . Logo, esse vetor é ortogonal a qualquer um dos vetores da matriz reduzida de autovetores U , resultado da decomposição espectral de XX^t . Isto é, o produto interno do centroide de coluna do espaço transformado com qualquer dos vetores de U é igual à zero. Portanto,

$$G = C F D_\sigma^{-1} - \left(\mathbf{1}_J \mathbf{m}_r^t D_r^{-\frac{1}{2}} \right) U$$

$$G = C F D_\sigma^{-1} - \mathbf{0}$$

$$G = C F D_\sigma^{-1}$$

Nesta fórmula de transição calcula-se o vetor de coordenadas G a partir do vetor de coordenadas F , isto é, a partir das coordenadas principais de linha são definidas as coordenadas principais de coluna.

Semelhantemente mostra-se que

$$F = R G D_\sigma^{-1}$$

em que as coordenadas principais de coluna definem as coordenadas principais de linha.

Na projeção conjunta das coordenadas principais de linha e de coluna, chamada de simétrica, as distâncias entre os perfis de linha e de coluna não são interpretáveis, mas apenas direções e tendências (SOUZA, 2004, p.86), pois pontos de linha e coluna muito próximos não quer dizer necessariamente uma forte associação entre eles (NYFJÄLL, 2002b, p. 7).

Uma alternativa a fim de tornar possível a comparação entre as nuvens de linha e coluna é padronizar as coordenadas principais de tal forma que as projeções tenham inércia igual a 1. Isso pode ser conseguido dividindo-se as componentes das projeções pela raiz da inércia sobre o eixo principal. Essas novas coordenadas chamam-se de coordenadas padronizadas e são mostradas a seguir:

Coordenadas padronizadas para linhas:

$$\Phi = FD_{\sigma}^{-1}$$

Coordenadas padronizadas para colunas:

$$\Gamma = GD_{\sigma}^{-1}$$

Assim como, foi necessário se calcular as coordenadas principais de linha em relação à coluna, devido a evitar problemas de sinais, faz-se aqui esta mesma necessidade (GREENACRE, 1984, p. 93), exceto se U foi calculado a partir de V ou vice-versa. Segue abaixo as fórmulas de transição entre coordenadas principais e padronizadas.

$$\Phi = FD_{\sigma}^{-1}$$

$$\Phi = RGD_{\sigma}^{-1}D_{\sigma}^{-1}$$

$$\Phi = RGD_{\sigma}^{-2}$$

$$\Gamma = GD_{\sigma}^{-1}$$

$$\Gamma = CFD_{\sigma}^{-1}D_{\sigma}^{-1}$$

$$\Gamma = CFD_{\sigma}^{-2}$$

Outra interpretação para as coordenadas padronizadas é que a mesma será a representação dos vértices nos eixos principais. Abaixo apresenta-se a evidencia desta relação.

As coordenadas do perfil de linha sobre os autovetores da matriz M obtida pela decomposição em valores singulares generalizada é dada por:

$$F = RGD_{\sigma}^{-1}$$

Portanto, GD_{σ}^{-1} é a matriz responsável por projetar os perfis de linha R sobre a base definida por M . Então, como os vértices da nuvem de linha representam as categorias de coluna e podem ser representados como vetores canônicos, pode-se substituir a matriz R por uma matriz identidade (representando uma base canônica). Logo percebe-se que esta projeção dará as coordenadas padronizadas de colunas, conforme apresentado abaixo

$$\Gamma = IGD_{\sigma}^{-1}$$

Como $G = CFD_{\sigma}^{-1}$, logo

$$\Gamma = CFD_{\sigma}^{-1}D_{\sigma}^{-1}$$

$$\Gamma = CFD_{\sigma}^{-2}$$

A mesma analogia pode ser realizada para a coordenada padronizada de linha.

Assim, as coordenadas padronizadas serão muito úteis na elaboração do mapa perceptual, pois representarão pontos de referência para análise.

Para o exemplo da tabela 2, pode-se decompor os resíduos dos perfis de linha $(R - \mathbf{1}_I \mathbf{m}_c^t)$ utilizando a decomposição em valores singulares generalizadas, conforme descrito a seguir:

$$(R - \mathbf{1}_I \mathbf{m}_c^t) = ND_{\sigma} M^t$$

em que,

$$N = D_r^{-\frac{1}{2}} U \quad \text{e} \quad M = D_c^{\frac{1}{2}} V$$

assim,

$$N = \begin{bmatrix} \frac{1}{\sqrt{0,15}} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{0,50}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{0,24}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{0,11}} \end{bmatrix} \begin{bmatrix} 0,017 & -0,165 \\ -0,647 & 0,076 \\ 0,609 & -0,536 \\ 0,459 & 0,824 \end{bmatrix} = \begin{bmatrix} 0,044 & -0,427 \\ -0,914 & 0,107 \\ 1,242 & -1,095 \\ 1,384 & 2,485 \end{bmatrix}$$

$$M = \begin{bmatrix} \sqrt{0,37} & 0 & 0 \\ 0 & \sqrt{0,36} & 0 \\ 0 & 0 & \sqrt{0,27} \end{bmatrix} \begin{bmatrix} 0,590 & 0,53 \\ -0,779 & 0,18 \\ 0,209 & -0,829 \end{bmatrix} = \begin{bmatrix} 0,359 & 0,322 \\ -0,468 & 0,108 \\ 0,108 & -0,430 \end{bmatrix}$$

portanto,

$$(R - \mathbf{1}_I \mathbf{m}_c^t) = ND_\sigma M^t$$

$$(R - \mathbf{1}_I \mathbf{m}_c^t) = \begin{bmatrix} 0,044 & -0,427 \\ -0,914 & 0,107 \\ 1,242 & -1,095 \\ 1,384 & 2,485 \end{bmatrix} \begin{bmatrix} 0,552 & 0 \\ 0 & 0,330 \end{bmatrix} \begin{bmatrix} 0,359 & -0,468 & 0,108 \\ 0,322 & 0,108 & -0,430 \end{bmatrix}$$

Para o exemplo da tabela 2, a Figura 25 ilustra a projeção dos vetores \mathbf{m}_1 e \mathbf{m}_2 da matriz M sob o espaço original.

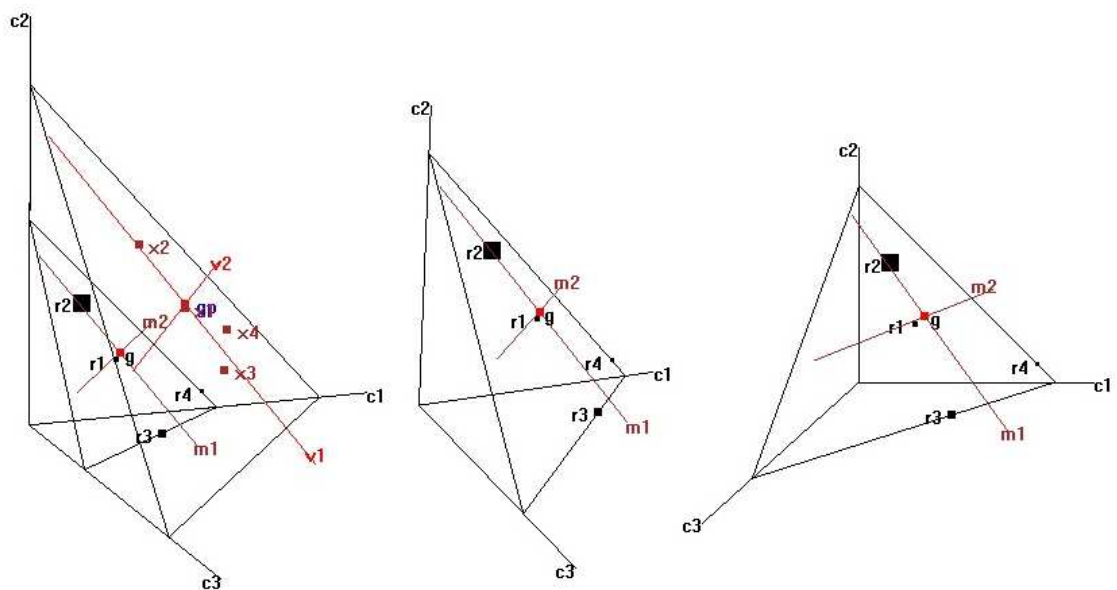


Figura 25 – Representação dos vetores \mathbf{m}_1 e \mathbf{m}_2 no espaço original

As coordenadas dos resíduos dos perfis de linha sobre os vetores m_1 e m_2 podem ser determinados através de:

$$F = ND_{\sigma}$$

$$F = \begin{bmatrix} 0,044 & -0,427 \\ -0,914 & 0,107 \\ 1,242 & -1,095 \\ 1,384 & 2,485 \end{bmatrix} \begin{bmatrix} 0,552 & 0 \\ 0 & 0,330 \end{bmatrix} = \begin{bmatrix} 0,024 & -0,141 \\ -0,505 & 0,035 \\ 0,686 & -0,361 \\ 0,764 & 0,820 \end{bmatrix}$$

A Figura 26 ilustra os vetores m_1 e m_2 formando uma base de um subespaço ótimo sob a métrica de qui-quadrado e as coordenadas dos resíduos dos perfis de linha $(R - \mathbf{1}_I m_c^t)$ sobre esta base é dada pela matriz F .

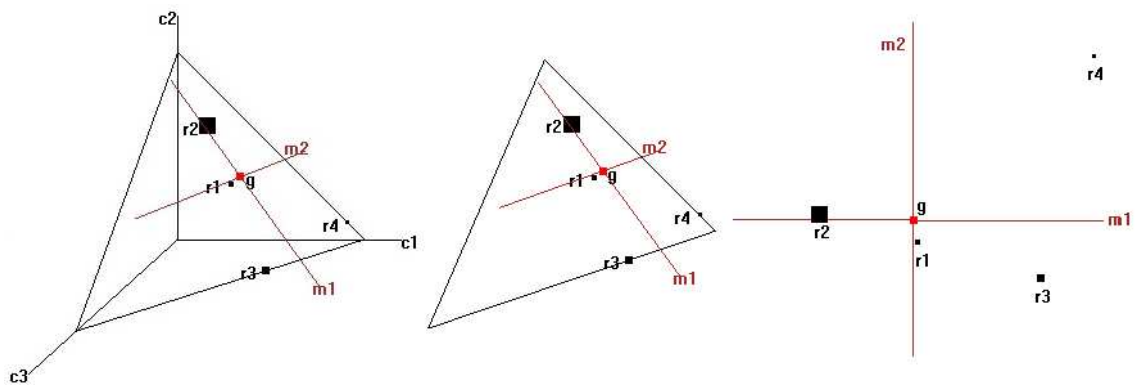


Figura 26 – Projeção dos resíduos dos perfis de linha $(R - \mathbf{1}_I m_c^t)$ sobre m_1 e m_2

As coordenadas dos resíduos dos perfis de coluna sobre os vetores l_1 e l_2 é dado por:

$$G = CFD_{\sigma}^{-1}$$

$$G = \begin{bmatrix} 0,14 & 0,27 & 0,32 & 0,27 \\ 0,14 & 0,83 & 0 & 0,03 \\ 0,19 & 0,37 & 0,44 & 0 \end{bmatrix} \begin{bmatrix} 0,024 & -0,141 \\ -0,505 & 0,035 \\ 0,686 & -0,361 \\ 0,764 & 0,820 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0,552 & 1 \\ 0 & 0,330 \end{bmatrix}$$

$$G = \begin{bmatrix} 0,536 & 0,288 \\ -0,717 & 0,099 \\ 0,222 & -0,526 \end{bmatrix}$$

As coordenadas padronizadas de linha e coluna serão respectivamente:

$$\Phi = RGD_{\sigma}^{-2}$$

$$\Phi = \begin{bmatrix} 0,33 & 0,33 & 0,33 \\ 0,20 & 0,60 & 0,20 \\ 0,50 & 0 & 0,50 \\ 0,91 & 0,09 & 0 \end{bmatrix} \begin{bmatrix} 0,536 & 0,288 \\ -0,717 & 0,099 \\ 0,222 & -0,526 \end{bmatrix} \begin{bmatrix} \frac{1}{0,552^2} & 0 \\ 0 & \frac{1}{0,330^2} \end{bmatrix}$$

$$\Phi = \begin{bmatrix} 0,044 & -0,427 \\ -0,914 & 0,107 \\ 1,242 & -1,095 \\ 1,384 & 2,485 \end{bmatrix}$$

$$\Gamma = CFD_{\sigma}^{-2}$$

$$\Gamma = \begin{bmatrix} 0,14 & 0,27 & 0,32 & 0,27 \\ 0,14 & 0,83 & 0 & 0,03 \\ 0,19 & 0,37 & 0,44 & 0 \end{bmatrix} \begin{bmatrix} 0,024 & -0,141 \\ -0,505 & 0,035 \\ 0,686 & -0,361 \\ 0,764 & 0,820 \end{bmatrix} \begin{bmatrix} \frac{1}{0,552^2} & 0 \\ 0 & \frac{1}{0,330^2} \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} 0,971 & 0,872 \\ -1,299 & 0,300 \\ 0,402 & -1,594 \end{bmatrix}$$

2.2.14 Pontos suplementares

Pontos suplementares são pontos que não tomam parte no desenvolvimento da AC, ou seja, não interferem na decomposição da matriz de correspondência, desejando-se desses pontos apenas suas projeções no mapa resultante da AC previamente realizada (SOUZA, 2004, p.72; NAITO, 2007, p.44).

Souza (2004, p.73) apresenta as origens mais comuns para os pontos suplementares: linhas ou colunas que distorcem a análise (*outliers*), observações adicionais ou novas informações em estudo. Ao representar o ponto suplementar no mapa perceptual, pode-se verificar as relações entre estes e as categorias estudadas previamente.

A determinação das coordenadas dos pontos suplementares segue a mesma ideia apresentada ao final do item anterior. Considerando-se GD_{σ}^{-1} a matriz responsável pela projeção dos pontos sobre a base definida por M e sendo r_s o perfil de linha do ponto suplementar, então

$$\mathbf{f}_s^t = \mathbf{r}_s^t G D_\sigma^{-1}$$

em que \mathbf{f}_s é o vetor de coordenadas da projeção do ponto suplementar em M .

Semelhantemente,

$$\mathbf{g}_s^t = \mathbf{c}_s^t F D_\sigma^{-1}$$

sendo \mathbf{g}_s o vetor de coordenadas da projeção do ponto suplementar r em L e \mathbf{c}_s o seu perfil de coluna.

2.2.15 Interpretação da AC

De nada adiantam cálculos complexos e exaustivos sobre uma técnica estatística se a mesma não for de fácil interpretação e útil para análise dos resultados. Assim, todo trabalho desenvolvido até agora envolvendo conceitos, demonstrações algébricas e visões geométricas dão subsídios para apresentação amigável dos relacionamentos e divergências existentes entre as categorias de uma tabela de contingência. Esta apresentação se dará através de um gráfico denominado mapa perceptual que será abordado adiante.

Apesar da AC ser uma técnica de cunho exploratório e da facilidade de interpretação dos mapas, há muitas controvérsias na comunidade científica especializada em relação a seus resultados (GREENACRE, 2007, p. 267; NAITO, 2007, p. 52-53), desde a forma de interpretação do mapa perceptual até a significância estatística das relações apresentadas. Porém, hoje já existem métodos que permitem investigar as propriedades estatísticas dos resultados da AC, incluindo testes de hipóteses, que dão a ela um caráter inferencial. (NAITO, 2007, p. 55). As críticas à interpretação dos resultados da AC servem como um alerta, lembrando que todo cuidado deve ser tomado, pois o uso apenas do mapa pode levar a uma interpretação duvidosa.

2.2.15.1 Medidas de auxílio na interpretação

Para uma interpretação mais acurada da AC faz-se necessário o uso de algumas medidas descritivas que auxiliam a interpretação do mapa

perceptual, tais como a decomposição da inércia sobre os eixos e as relações entre os pontos e os eixos (NAITO, 2007, p. 48).

Proporção de inércia sobre os eixos

A inércia de cada uma das nuvens $N(I)$ e $N(J)$ é decomposta sobre seus j e i eixos principais, já mencionado no item 2.2.11, e como os eixos que não captam alguma inércia são descartados, pode-se decompor a inércia da nuvem em k eixos principais em que $k = \min(i - 1, j - 1)$ e a inércia captada pelo eixo α é dado por λ_α . Então,

$$In(I) = In(J) = tr(D_\lambda) = \lambda_1 + \lambda_2 + \dots + \lambda_\alpha + \dots + \lambda_k$$

Portanto, pode-se medir a importância relativa de cada eixo para reter a inércia total da nuvem sobre si mesmo e esta medida é dada pela seguinte proporção:

$$\omega_\alpha = \frac{\lambda_\alpha}{\sum_{\alpha=1}^k \lambda_\alpha} \cdot 100\%$$

Para redução de dimensionalidade selecionam-se os principais eixos que captam a inércia da nuvem de pontos, e que ao mesmo tempo possam ser representados de forma bi ou tridimensional. Sendo assim, pode-se medir o quão boa é esta representação, através do somatório dos ω_α dos eixos envolvidos na representação. Deve-se procurar manter um número de eixos tal que a porcentagem de inércia explicada seja superior a 50% (NAITO, 2007, p. 49).

Contribuição absoluta

A contribuição absoluta é o quanto um perfil (ponto) contribui percentualmente para a formação da inércia total do eixo, que por consequência influencia na direção do eixo. Portanto, aqueles perfis que possuem maior contribuição absoluta e que estão em oposição em relação à origem no mapa perceptual serão utilizados para dar uma interpretação do

significado do eixo, normalmente fazendo uso de oposição de adjetivos (PAMPLONA, 1998, p. 43).

A contribuição absoluta é uma medida dada pela razão entre a inércia do ponto ou projetada no eixo e a inércia total captada pelo eixo. Considerando um perfil de linha, o eixo, a massa dada por, e a componente da matriz de coordenadas principais, tem-se que:

$$\text{---} \quad \text{---}$$

Para um perfil de coluna e o eixo correspondente, tem-se que

$$\text{---} \quad \text{---}$$

Contribuição relativa

Conforme já mencionado no item 2.2.11 a inércia de um perfil pode ser decomposta nos eixos principais. A Figura 27 ilustra a relação das componentes da matriz de coordenadas principais e a inércia do ponto.

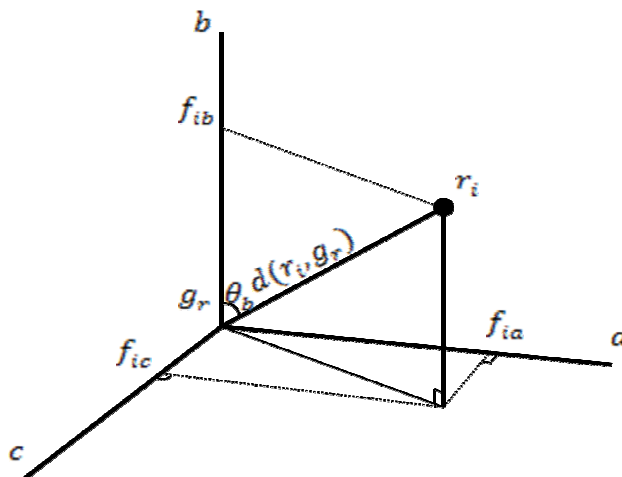


Figura 27 – Decomposição da distância do perfil de linha pertencente ao sobre os eixos a, b, c

Pelo teorema de Pitágoras, tem-se que

$$d^2(r_i, g_r) = \sum_{\alpha=1}^k f_{i\alpha}^2$$

Multiplicando cada membro pela massa do ponto, tem-se

$$m_{r_i} d^2(r_i, g_r) = m_{r_i} \sum_{\alpha=1}^k f_{i\alpha}^2$$

$$In(r_i) = \sum_{\alpha=1}^k In(r_i)_\alpha$$

A contribuição relativa representa o quanto um perfil de linha é representado pelo eixo. Dado um perfil de linha r_i e o eixo k , tem-se:

$$cr(\alpha)_{r_i} = \frac{In(r_i)_\alpha}{In(r_i)} = \frac{In(r_i)_\alpha}{\sum_{\alpha=1}^k In(r_i)_\alpha} = \frac{m_{r_i} f_{i\alpha}^2}{m_{r_i} \sum_{\alpha=1}^k f_{i\alpha}^2} = \frac{f_{i\alpha}^2}{d^2(r_i, g_r)} = \cos^2 \theta_{r_i\alpha}$$

$$cr(\alpha)_{c_j} = \frac{In(c_j)_\alpha}{In(c_j)} = \frac{In(c_j)_\alpha}{\sum_{\alpha=1}^k In(c_j)_\alpha} = \frac{m_{c_j} g_{j\alpha}^2}{m_{c_j} \sum_{\alpha=1}^k g_{j\alpha}^2} = \frac{g_{j\alpha}^2}{\sum_{\alpha=1}^k g_{j\alpha}^2} = \cos^2 \theta_{c_j\alpha}$$

A contribuição relativa é que auxilia na escolha da dimensão que melhor caracteriza um ponto, isto é, quanto maior a contribuição relativa de um eixo, mais esse eixo caracteriza o ponto (PAMPLONA, 1998, p. 42).

2.2.15.2 Mapas perceptuais

Os mapas perceptuais permitem visualizar as similaridades como também as diferenças entre as categorias das variáveis envolvidas. O tipo de mapa a ser utilizado depende da natureza dos dados e dos objetivos do pesquisador (GREENACRE, 2007, p.11).

No item 2.2.13, obteve-se as coordenadas dos pontos das nuvens por dois processos diferentes: por coordenadas principais e por coordenadas padronizadas. A conjugação desses processos na representação dos perfis de

linha e coluna simultaneamente gera dois tipos básicos de mapas, os simétricos e os assimétricos.

Não importando o tipo de mapa, os resultados da AC são invariantes (GREENACRE, 2007, p. 11). Pamplona (1998, p. 42-45) apresenta um roteiro geral de interpretação dos mapas.

A primeira etapa da interpretação do mapa perceptual é dar um significado aos eixos, de acordo com os pontos com as maiores contribuições absolutas e em oposição à origem, interpretando os eixos por oposição de adjetivos.

Para o estudo dos pontos individualmente deve-se levar em conta a contribuição relativa de cada eixo para o ponto, compará-lo com a média (centroide, origem), e verificar sua posição em relação ao significado dos eixos.

A relação entre pontos de uma mesma nuvem pode ser dada por proximidades ou por distanciamentos. Pontos muito próximos do centroide estão em situação de independência, ainda que bastante agrupados. Pontos de uma mesma nuvem muito próximos entre si, mas distanciados do centroide, apresentam estruturas semelhantes. Pontos distantes entre si apresentam relação com categorias diferentes da outra variável.

Mapa simétrico

O mapa simétrico é aquele que utiliza as coordenadas principais definidas em F para os perfis de linha e as coordenadas principais definidas em G para os perfis de coluna sobre o mesmo gráfico (GREENACRE, 2007, p. 70).

Esse gráfico é um dos mais utilizados em AC (GREENACRE, 2007, p.70), pois permite interpretar tanto a projeção dos pontos de linha como dos pontos de coluna em relação aos eixos num único gráfico, sendo que quanto mais próximo da origem menor é a contribuição relativa desse eixo para com o ponto e quanto mais à projeção se aproxima de uma das extremidades do eixo, maior é a associação desse ponto com o significado da extremidade do eixo.

Outra grande vantagem desse mapa é que as coordenadas dos pontos retrata a distância de qui-quadrado projetada sobre o eixo correspondente. Assim, pode-se verificar as relações de similaridades através das distâncias entre os pontos da mesma nuvem. É importante ressaltar que como os perfis de linha e coluna não estão no mesmo subespaço, não é possível afirmar

relações de similaridades ou divergências entre perfis de linha e coluna baseados nas suas distâncias. Tais relações podem ser conseguidas através dos mapas assimétricos.

Mapas assimétricos

Os mapas assimétricos são aqueles que utilizam coordenadas principais para uma nuvem e padronizadas para a outra nuvem. Quando utiliza coordenada principal por linha (matriz F) e padronizada por coluna (matriz Γ), é dito mapa principal por linha; quando utiliza coordenada principal por coluna (matriz G) e padronizada por linha (matriz Φ), é dito mapa principal por coluna (GREENACRE, 2007, p. 68).

Conforme mencionado no item 2.2.13 as coordenadas padronizadas de linha e coluna, podem ser interpretadas como vértices na nuvem de coluna e de linha, respectivamente. Sendo assim, os perfis de linha e os vértices de coluna estão no mesmo subespaço e, portanto, podem ser identificadas suas similaridades e diferenças através de suas distâncias.

Se o objetivo é estudar o comportamento das linhas em relação às colunas, utiliza-se o mapa principal por linha, isto é, linhas em coordenadas principais e colunas em coordenadas padronizadas. Este gráfico irá representar os perfis de coluna como vértices (valores extremos) e assim verifica-se para cada perfil de linha de qual vértice mais se aproximou.

Se o objetivo é estudar o comportamento das colunas em relação às linhas, utiliza-se o mapa principal por coluna, isto é, colunas em coordenadas principais e linhas em coordenadas padronizadas, visto que as linhas serão os vértices (valores extremos).

A seguir ilustra-se a interpretação da AC para os dados da tabela 2. Primeiramente, serão calculadas as medidas que auxiliam na interpretação dos mapas perceptuais para posteriormente construí-los.

Sabe-se que

$$D_{\sigma}^2 = D_{\lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 0,305 & 0 \\ 0 & 0,109 \end{bmatrix}$$

portanto, calcula-se a proporção da inércia da nuvem que é explicada por cada eixo (1 e 2).

$$\omega_1 = \frac{\lambda_1}{\sum_{\alpha=1}^2 \lambda_{\alpha}} \cdot 100\% = \frac{0,305}{0,414} \cdot 100\% = 73,7\%$$

$$\omega_2 = \frac{\lambda_2}{\sum_{\alpha=1}^2 \lambda_{\alpha}} \cdot 100\% = \frac{0,109}{0,414} \cdot 100\% = 26,3\%$$

O eixo 1 é responsável por 73,7% da inércia total, sendo o eixo 2 responsável por 26,3%. Tem-se que 100% da inércia total está capturada por esses dois eixos.

Supondo que o objetivo seja analisar a relação das linhas em detrimento das colunas. Então as maiores contribuições absolutas dos perfis de coluna darão o significado dos eixos, permitindo assim, uma interpretação dos perfis de linha sobre os eixos. Calculando a contribuição absoluta dos perfis de coluna sobre os eixos 1 e 2, tem-se:

Sabe-se que:

$$\Gamma = \begin{bmatrix} 0,971 & 0,872 \\ -1,299 & 0,300 \\ 0,402 & -1,594 \end{bmatrix}$$

Calculando as contribuições absolutas das colunas sobre o eixo 1:

$$ca(c_1)_1 = \frac{In(c_1)_1}{\sum_{j=1}^3 In(c_1)_1} = \frac{m_{c_1} g_{i1}^2}{\lambda_1} = m_{c_1} \tau_{11}^2 = 0,37 \cdot 0,971^2 = 0,349$$

$$ca(c_2)_1 = m_{c_2} \tau_{21}^2 = 0,36 \cdot 1,299^2 = 0,608$$

$$ca(c_3)_1 = m_{c_3} \tau_{31}^2 = 0,27 \cdot 0,402^2 = 0,044$$

Calculando as contribuições absolutas das colunas sobre o eixo 2:

$$ca(c_1)_2 = m_{c_1} \tau_{12}^2 = 0,37 \cdot 0,872^2 = 0,281$$

$$ca(c_2)_2 = m_{c_2} \tau_{22}^2 = 0,36 \cdot 0,300^2 = 0,032$$

$$ca(c_3)_2 = m_{c_3} \tau_{32}^2 = 0,27 \cdot (-1,594)^2 = 0,686$$

Sendo assim, são os perfis de coluna com as maiores contribuições absolutas e opostos ao centroide que irão dar significado aos extremos dos

eixos. Logo, o eixo 1 pode ser interpretado como a diferença característica entre c_1 e c_2 (equivalente às categorias B1 e B2 referente ao exemplo ilustrativo da tabela 2) e o eixo 2 pode ser interpretado como a diferença da característica entre c_1 e c_3 (categoria B1 e B3). Como as categorias de linha e coluna não têm significado neste exemplo, não é possível dar um significado a interpretação dos eixos.

Se o interesse for analisar as relações das colunas em relação às linhas. Faz-se necessário dar um significado aos eixos em relação às linhas e assim devem-se calcular as contribuições absolutas em relação às linhas

Sabe-se que:

$$\Phi = \begin{bmatrix} 0,044 & -0,427 \\ -0,914 & 0,107 \\ 1,242 & -1,095 \\ 1,384 & 2,485 \end{bmatrix}$$

Calculando as contribuições absolutas das linhas sobre o eixo 1:

$$ca(r_1)_1 = \frac{In(r_1)_1}{\sum_{i=1}^4 In(r_i)_1} = \frac{m_{r_1} f_{11}^2}{\lambda_1} = m_{r_1} \phi_{11}^2 = 0,15 \cdot 0,044^2 = 0,000$$

$$ca(r_2)_1 = m_{r_2} \phi_{21}^2 = 0,50 \cdot (-0,914)^2 = 0,418$$

$$ca(r_3)_1 = m_{r_3} \phi_{31}^2 = 0,24 \cdot (1,242)^2 = 0,370$$

$$ca(r_4)_1 = m_{r_4} \phi_{41}^2 = 0,11 \cdot (1,384)^2 = 0,211$$

Calculando as contribuições absolutas das linhas sobre o eixo 2:

$$ca(r_1)_2 = m_{r_1} \phi_{12}^2 = 0,15 \cdot (-0,427)^2 = 0,027$$

$$ca(r_2)_2 = m_{r_2} \phi_{22}^2 = 0,50 \cdot (0,107)^2 = 0,006$$

$$ca(r_3)_2 = m_{r_3} \phi_{32}^2 = 0,24 \cdot (-1,095)^2 = 0,288$$

$$ca(r_4)_2 = m_{r_4} \phi_{42}^2 = 0,11 \cdot (2,485)^2 = 0,679$$

A seguir calculam-se as contribuições relativas referente dos perfis de linha.

Sabe-se que a diagonal da matriz XX^t revela a inércia dos perfis de linha,

$$XX^t = \begin{bmatrix} 0,003 & -0,005 & 0,013 & -0,012 \\ -0,005 & 0,128 & -0,124 & -0,083 \\ 0,013 & -0,124 & 0,144 & 0,037 \\ -0,012 & -0,08 & 0,371 & 0,138 \end{bmatrix}$$

e que a matriz F fornece as coordenadas principais de linha

$$F = \begin{bmatrix} 0,024 & -0,141 \\ -0,505 & 0,035 \\ 0,686 & -0,361 \\ 0,764 & 0,820 \end{bmatrix}$$

Calculando as contribuições relativas do eixo 1 sobre os perfis de linha:

$$cr(1)_{r_1} = \frac{In(r_1)_1}{In(r_1)} = \frac{m_{r_1}f_{11}^2}{In(r_1)} = \frac{0,15 \cdot 0,024^2}{0,003} = 0,029$$

$$cr(1)_{r_2} = \frac{In(r_2)_1}{In(r_2)} = \frac{m_{r_2}f_{21}^2}{In(r_1)} = \frac{0,50 \cdot (-0,505)^2}{0,128} = 0,995$$

$$cr(1)_{r_3} = \frac{In(r_3)_1}{In(r_3)} = \frac{m_{r_3}f_{31}^2}{In(r_3)} = \frac{0,24 \cdot (0,686)^2}{0,144} = 0,783$$

$$cr(1)_{r_4} = \frac{In(r_4)_1}{In(r_4)} = \frac{m_{r_4}f_{41}^2}{In(r_4)} = \frac{0,11 \cdot (0,764)^2}{0,138} = 0,464$$

Calculando as contribuições relativas do eixo 2 sobre os perfis de linha:

$$cr(2)_{r_1} = \frac{In(r_1)_2}{In(r_1)} = \frac{m_{r_1}f_{12}^2}{In(r_1)} = \frac{0,15 \cdot (-0,141)^2}{0,003} = 0,971$$

$$cr(2)_{r_2} = \frac{In(r_2)_2}{In(r_2)} = \frac{m_{r_2}f_{22}^2}{In(r_1)} = \frac{0,50 \cdot 0,035^2}{0,128} = 0,005$$

$$cr(2)_{r_3} = \frac{In(r_3)_2}{In(r_3)} = \frac{m_{r_3}f_{32}^2}{In(r_3)} = \frac{0,24 \cdot (-0,361)^2}{0,144} = 0,217$$

$$cr(2)_{r_4} = \frac{In(r_4)_2}{In(r_4)} = \frac{m_{r_4}f_{42}^2}{In(r_4)} = \frac{0,11 \cdot 0,82^2}{0,138} = 0,536$$

Pelas contribuições relativas pode-se perceber que a análise do perfil de linha r_1 deve ser realizada sobre o eixo 2, visto que o mesmo contribui com

97,1% da inércia deste perfil, enquanto o eixo 1 contribui com apenas 2,9% da inércia, podendo-se, portanto, ignorá-lo. Já a análise da categoria A4, referente ao perfil de linha r_4 , deve-se levar em consideração os dois eixos, visto que a contribuição relativa do eixo 1 e 2 são 46,4% e 53,6%, respectivamente.

A seguir calculam-se as contribuições relativas referentes aos perfis de coluna.

Sabe-se que a diagonal da matriz X^tX revela a inércia dos perfis de coluna,

$$X^tX = \begin{bmatrix} 0,137 & -0,13 & -0,01 \\ -0,13 & 0,189 & -0,066 \\ -0,01 & -0,066 & 0,088 \end{bmatrix}$$

e que a matriz G fornece as coordenadas principais de coluna

$$G = \begin{bmatrix} 0,536 & 0,288 \\ -0,717 & 0,099 \\ 0,222 & -0,526 \end{bmatrix}$$

Calculando as contribuições relativas do eixo 1 sobre os perfis de coluna:

$$cr(1)_{c_1} = \frac{In(c_1)_1}{In(c_1)} = \frac{m_{c_1}g_{11}^2}{In(c_1)} = \frac{0,37 \cdot 0,536^2}{0,137} = 0,776$$

$$cr(1)_{c_2} = \frac{In(c_2)_1}{In(c_2)} = \frac{m_{c_2}g_{21}^2}{In(c_2)} = \frac{0,36 \cdot (-0,717)^2}{0,189} = 0,981$$

$$cr(1)_{c_3} = \frac{In(c_3)_1}{In(c_3)} = \frac{m_{c_3}g_{31}^2}{In(c_3)} = \frac{0,27 \cdot 0,222^2}{0,088} = 0,151$$

Calculando as contribuições relativas do eixo 2 sobre os perfis de coluna:

$$cr(2)_{c_1} = \frac{In(c_1)_2}{In(c_1)} = \frac{m_{c_1}g_{12}^2}{In(c_1)} = \frac{0,37 \cdot 0,288^2}{0,137} = 0,224$$

$$cr(2)_{c_2} = \frac{In(c_2)_2}{In(c_2)} = \frac{m_{c_2}g_{22}^2}{In(c_2)} = \frac{0,36 \cdot 0,099^2}{0,189} = 0,019$$

$$cr(2)_{c_3} = \frac{In(c_3)_2}{In(c_3)} = \frac{m_{c_3} g_{32}^2}{In(c_3)} = \frac{0,27 \cdot (-0,526)^2}{0,088} = 0,849$$

De forma similar como foi realizado para os perfis de linha, pode-se perceber que a categoria B2, referente ao perfil de coluna c_2 , deve ser analisada levando em consideração apenas o eixo 1, visto que o mesmo é responsável por 98,1% da inércia dessa categoria.

Finalmente podem-se construir os mapas perceptuais para a análise das similaridades e divergências entre as categorias do exemplo ilustrativo da tabela 2. Através das matrizes F e G que fornecem as coordenadas principais da nuvem de linha e coluna, respectivamente pode-se montar o mapa simétrico conforme descrito a seguir:

Sabe-se que,

$$F = \begin{bmatrix} 0,024 & -0,141 \\ -0,505 & 0,035 \\ 0,686 & -0,361 \\ 0,764 & 0,820 \end{bmatrix} \quad \text{e} \quad G = \begin{bmatrix} 0,536 & 0,288 \\ -0,717 & 0,099 \\ 0,222 & -0,526 \end{bmatrix}$$

então,

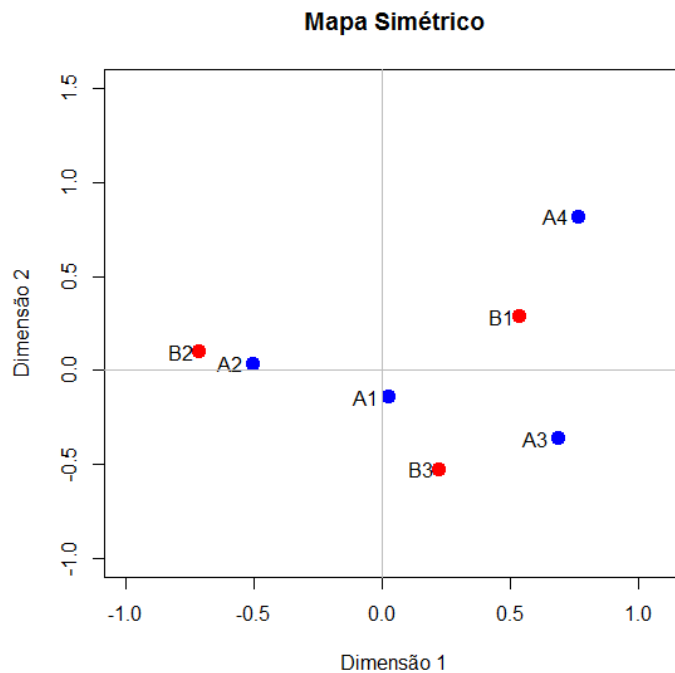


Figura 28 – Mapa perceptual simétrico do exemplo ilustrativo da Tabela 2

Com o auxílio das medidas descritivas (proporção da inércia de cada eixo, contribuições absolutas e relativas), pode-se realizar a interpretação do mapa perceptual simétrico.

Através da proporção da inércia que é captada por cada eixo, percebe-se que a dimensão 1 (eixo 1) explica 73,7% de toda a inércia da nuvem, sendo o restante explicado pela segunda dimensão (eixo 2). De forma geral a dimensão 1 será a mais importante para análise dos dados.

Com o objetivo da análise sobre as linhas, pode-se interpretar a dimensão 1 através das maiores contribuições absolutas de coluna opostas no mapa. Assim, B1 e B2 darão o significado da dimensão 1 e B1 e B3 darão o significado da dimensão 2.

Na análise gráfica pode-se perceber que A1 está próxima ao centroide e portanto é a categoria menos relacionada com as categorias B_j . Já a categoria A2 deve ser analisada levando-se em consideração a dimensão 1, pois a contribuição relativa do eixo 1 para sua inércia é de 99,5%.

Vale a pena ressaltar que nesse tipo de gráfico não é possível avaliar proximidades e diferenças entre categorias de linha e coluna, visto que as mesmas não estão no mesmo subespaço. Assim, analisando o mapa simétrico pode-se ter a impressão que a categoria B1 possui a mesma distância entre as categorias A3 e A4, porém através do mapa assimétrico principal por linha (Figura 29) é possível perceber que a categoria B1 está próxima de A4, mas não de A3, o que sugere uma relação entre B1 e A4.

O mapa assimétrico é utilizado quando se deseja relacionar as categorias de linha e coluna. No mapa assimétrico principal por linha utilizam-se as matrizes F e Γ que revela as coordenadas principais por linha e padronizadas por coluna respectivamente.

Sabe-se que

$$F = \begin{bmatrix} 0,024 & -0,141 \\ -0,505 & 0,035 \\ 0,686 & -0,361 \\ 0,764 & 0,820 \end{bmatrix} \quad \text{e} \quad \Gamma = \begin{bmatrix} 0,971 & 0,872 \\ -1,299 & 0,300 \\ 0,402 & -1,594 \end{bmatrix}$$

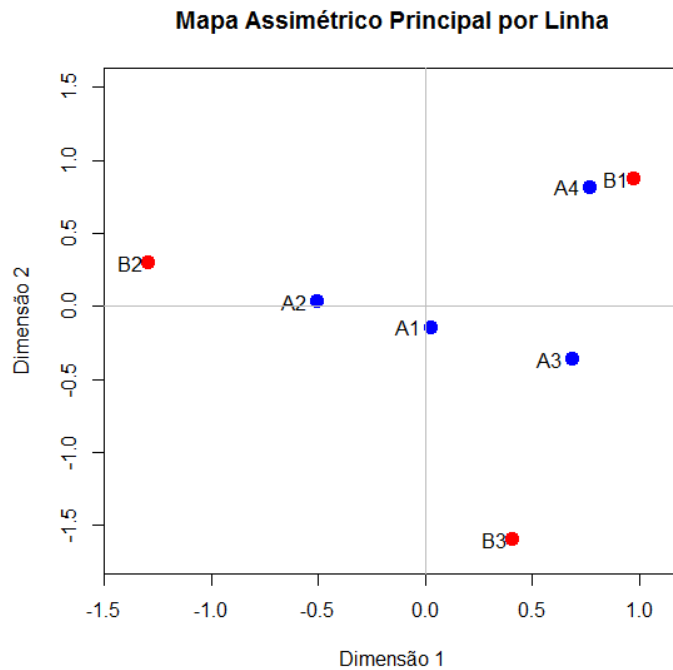


Figura 29 – Mapa perceptual assimétrico principal por linha do exemplo ilustrativo da Tabela 2

Utilizando o mapa assimétrico por linha, os perfis de coluna (categorias B) serão os pontos extremos que os perfis de linha (categorias A) podem tomar. Assim, quanto mais próximo do centroide uma categoria A estiver, mais independente estará em relação às categorias de B e por sua vez quanto mais próxima uma categoria A estiver de uma das categorias de B, mais é evidente sua relação com a mesma. Nesse, sentido percebe-se a forte relação que há entre A4 e B1.

A análise do mapa assimétrico principal por coluna (Figura 30) pode ser realizada da mesma forma que o mapa assimétrico principal por linha. Para montagem deste mapa utilizam-se as matrizes G e Φ que constituem as coordenadas principal por coluna e padronizadas por linha.

Sabe-se que

$$\Phi = \begin{bmatrix} 0,044 & -0,427 \\ -0,914 & 0,107 \\ 1,242 & -1,095 \\ 1,384 & 2,485 \end{bmatrix} \quad \text{e} \quad G = \begin{bmatrix} 0,536 & 0,288 \\ -0,717 & 0,099 \\ 0,222 & -0,526 \end{bmatrix}$$

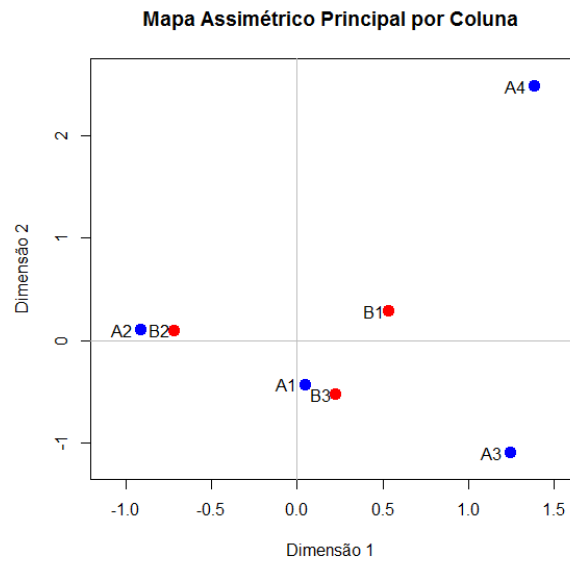


Figura 30 – Mapa perceptual assimétrico principal por coluna do exemplo ilustrativo da Tabela 2

3 MATERIAL E MÉTODOS

A produção de cana-de-açúcar (*Saccharum spp.*) na safra de 2011/2012, estimada pela Conab (2011) foi de 571 milhões de toneladas e de acordo com Bertelli (2012) esta produção necessitará ser dobrada até 2020, para atender à fabricação de açúcar para o mercado doméstico e exportações, a produção de etanol, bem como na geração de 13 mil megawatts, o que equivale à produção da hidrelétrica de Itaipu. A cana-de-açúcar é uma fonte de energia renovável e produzida a partir de seus resíduos, tais como bagaço, palha, palhiço, entre outros (ALCARDE, 2012)

Para satisfazer esta demanda de mercado fazem-se necessários investimentos nos diversos setores da economia envolvidos na cadeia produtiva da cana-de-açúcar, tais como aumento das áreas de plantio, oferta de crédito para o aumento da produção e o desenvolvimento de novas tecnologias, tanto para o manejo da cana, quanto para o seu melhoramento genético.

Os estudos de melhoramento genético envolvendo a cana-de-açúcar visam obter o aumento de produtividade e melhoria na qualidade da matéria prima para fabricação de etanol e de açúcar. Como o desenvolvimento de novos cultivares é um processo de longa duração, torna-se indispensável o uso de técnicas estatísticas ou fitotécnicas apropriadas para se otimizar o processo de seleção.

Uma alternativa para agilizar a coleta de dados no campo é a categorização das variáveis componentes da produção, pois desta forma não requer avaliação individual de plantas no experimento. Apesar de ser um processo menos eficiente de seleção, a categorização permite a avaliação de um maior número de famílias por vez (SOUZA; PETERNELLI; BERNARDES, 2012, p. 2).

Em 2006, a Estação Experimental de Serra do Ouro, localizada na cidade de Murici (AL) e pertencente à Universidade Federal de Alagoas, realizou vários cruzamentos de cana-de-açúcar e os enviou ao Centro de Pesquisa e Melhoramento de Cana-de-Açúcar da Universidade Federal de Viçosa (CECA - UFV), que por sua vez realizou 5 experimentos com 22 famílias de irmãos completos cada um.

Para a aplicação da AC foi utilizada uma base de dados proveniente do experimento 3. Este era formado por 22 famílias em blocos inteiramente casualizados com 5 repetições. As variáveis usadas no presente trabalho foram o número de colmos por planta (NC), a altura média dos colmos (AC), o diâmetro do 5º internódio contado a partir da base do colmo (DC) e o peso da touceira.

Com base em critérios de prática de campo, os dados foram categorizados conforme apresentado na Tabela 6:

Tabela 6 – Critério de categorização das variáveis explicativas do experimento com cana-de-açúcar

| Variável | Categoria | Medida |
|------------------------|--------------------|---|
| Número de colmos (NC) | pouco (po) | $NC < 5$ |
| | médio (me) | $5 \leq NC < 8$ |
| | muitos (mu) | $NC \geq 8$ |
| Altura do colmo (AC) | baixo (ba) | $AC < 260 \text{ cm}$ |
| | alto (al) | $AC \geq 260 \text{ cm}$ |
| Diâmetro do colmo (DC) | fino (fin) | $DC \leq 2 \text{ cm}$ |
| | médio (med) | $2 \text{ cm} < DC \leq 2,5 \text{ cm}$ |
| | grosso (gro) | $2,5 \text{ cm} < DC \leq 3 \text{ cm}$ |
| | muito grosso (mgr) | $DC > 3 \text{ cm}$ |

Fonte: Adaptado de Souza, Peternelli e Bernardes, 2012, p.3.

O peso das touceiras (em TCH – tonelada de cana por hectare) foi categorizado para este trabalho utilizando, como ponto de corte, os percentis 20, 40, 60 e 80, tendo como resultado: Péssimo Peso, se peso $\leq 2,8$; Peso Ruim, se $2 < \text{peso} \leq 6,4$; Peso Regular, se $6,4 < \text{peso} \leq 11,2$; Bom Peso, se $11,2 < \text{peso} \leq 18,4$; e Excelente Peso, se peso $> 18,4$.

As análises foram realizadas no *software* R versão 2.14.0 (R DEVELOPMENT CORE TEAM, 2011) por meio do pacote *ca* (GREENACRE, NENADIĆ, 2010), com mapas perceptuais principal por linha e simétrico. Posteriormente foram adicionados pontos suplementares por adição de novas informações referentes às famílias. Além disso, foi consultada a obra de Peternelli e Mello (2011) para lidar com os aspectos operacionais do *software* R.

O uso da AC na área de ciências agrárias ainda não é muito explorado. Com este trabalho, além do detalhamento do desenvolvimento algébrico e geométrico já apresentado, procura-se demonstrar sua utilidade aplicando-a aos dados do experimento de melhoramento genético da cana-de-açúcar. Neste exemplo, objetivou-se definir qual característica fenotípica da cana, associada aos componentes da produção, que pode ser utilizada como indicadora do melhor peso da touceira, como também definir quais são as famílias que apresentaram melhores resultados.

4 RESULTADOS E DISCUSSÃO

Para facilitar o entendimento da interpretação da AC, tanto em relação às medidas descritivas dos perfis, quanto para análise dos mapas perceptuais buscou-se uma nomenclatura para as combinações entre as categorias das variáveis, utilizando as abreviações apresentadas na Tabela 6 na seguinte ordem: número de colmos, altura e diâmetro. Exemplificando, mealfin quer dizer uma touceira com as características de número médio de colmos, altos e finos; e mubamgr, muitos colmos, baixos e muito grossos.

Os pesos foram classificados em p_pess (péssimo peso), p_ruim (peso ruim), p_reg (peso regular), p_bom (bom peso) e p_exc (excelente peso).

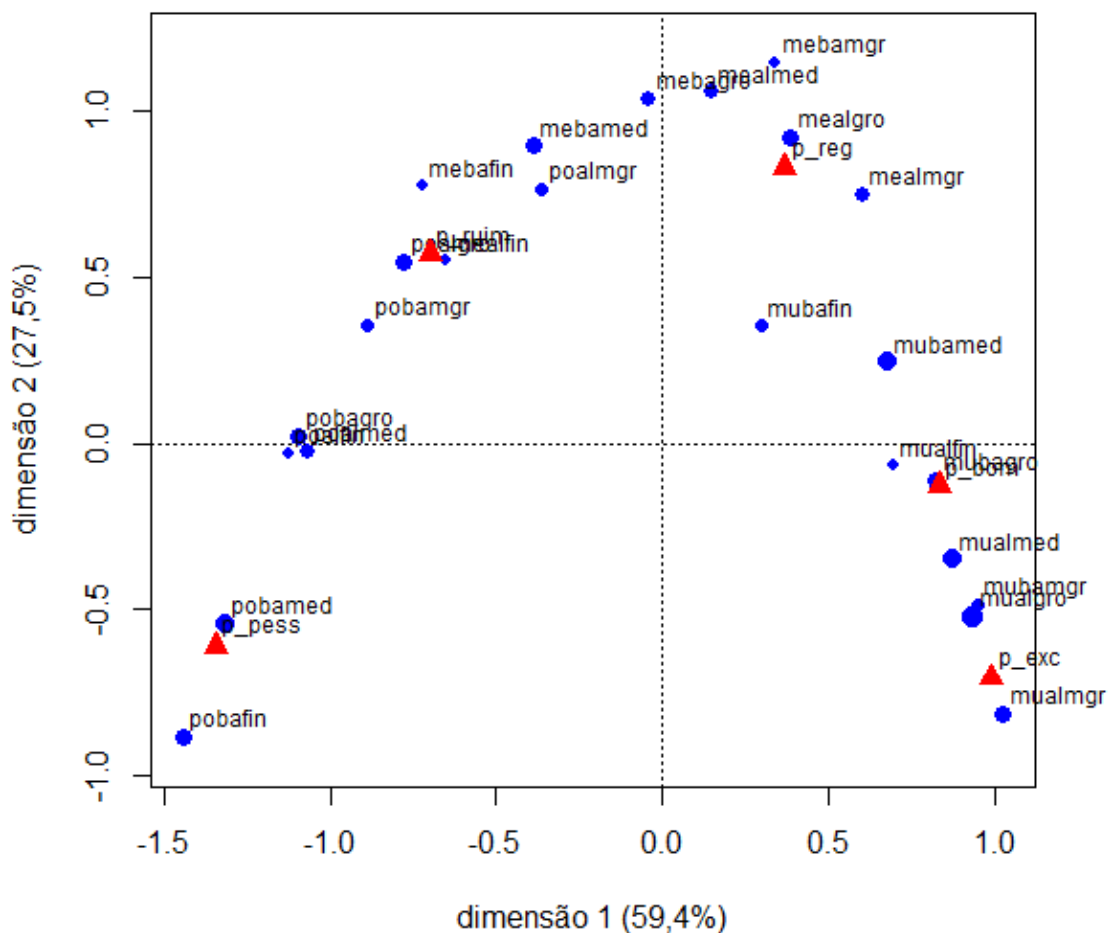


Figura 31 – Mapa perceptual simétrico das características da cana-de-açúcar e peso da cana

O mapa perceptual simétrico (Figura 31) apresenta uma conformação característica chamada de ferradura (*horseshoe*) ou efeito Gutmann. Esta conformação pode aparecer em AC e também em outras técnicas de escalonamento multidimensional (GREENACRE, 1984, p. 227), e, segundo Naito (2007, p. 53), não compromete os resultados da AC para o conjunto de dados, além de permitir, afirma Greenacre (1984, p.227), uma reordenação dos pontos como aparecem em relação ao eixo de maior inércia. Tabela 7 – Decomposição da inércia sobre os eixos principais.

Tabela 7 – Decomposição da inércia sobre os eixos principais

| Dimensão | Valor | % | Cum% |
|----------|----------|-------|-------|
| 1 | 0,827438 | 59,4 | 59,4 |
| 2 | 0,383060 | 27,5 | 86,9 |
| 3 | 0,156063 | 11,2 | 98,1 |
| 4 | 0,026735 | 1,9 | 100,0 |
| Total: | 1,39330 | 100,0 | |

Pela Tabela 7, verificou-se que o eixo 1 é responsável por 59,4% da inércia total e o eixo 2 por 27,5%, que conjuntamente perfazem 86,9% da inércia total. Logo, existe boa representatividade dos perfis no espaço \mathbb{R}^2 .

Como o objetivo é analisar as características fenotípicas da cana-de-açúcar em relação ao peso da touceira, buscou-se dar significado aos eixos levando-se em consideração os perfis de coluna que representam a categorização do peso.

Tabela 8 – Medidas descritivas da AC sobre o perfil de coluna multiplicadas por mil.

| Perfil | mass (1) | qit (2) | inr (3) | k=1 (4) | cor (5) | ctr (6) | k=2 (7) | cor (8) | ctr (9) |
|--------|-------------|------------|------------|------------|------------|------------|------------|------------|------------|
| p_pess | 210 | 959 | 342 | -1345 | 795 | 459 | -610 | 163 | 203 |
| p_ruim | 204 | 687 | 173 | -697 | 411 | 120 | 571 | 276 | 174 |
| p_reg | 205 | 808 | 150 | 369 | 134 | 34 | 830 | 674 | 368 |
| p_bom | 191 | 859 | 114 | 835 | 840 | 161 | -127 | 19 | 8 |
| p_exc | 191 | 918 | 221 | 992 | 610 | 227 | -704 | 308 | 247 |

(1) – massa do perfil

(2) – contribuição relativa total considerando as dimensões 1 e 2

(3) – inércia relativa

(4) – coordenada principal sobre o eixo 1

(5) – contribuição relativa sobre o eixo 1

- (6) – contribuição absoluta sobre o eixo 1 (8) – contribuição relativa sobre o eixo 2
(7) – coordenada principal sobre o eixo 2 (9) – contribuição absoluta sobre o eixo 2

Pela Tabela 8 percebe-se que os perfis de coluna p_pess e p_exc foram os que mais contribuíram para a inércia e direção do eixo 1, com contribuições absolutas de 45,9% e 22,7%, respectivamente, e como esses perfis se apresentam em lados opostos no mapa perceptual serão utilizados para dar um significado ao eixo 1. Assim, quanto mais os perfis de linha se aproximam da extremidade esquerda do eixo, mais pode-se associar esses perfis a um peso baixo da touceira, e quanto mais os perfis se aproximam da extremidade direita do eixo mais pode-se associar esses perfis a um peso alto da touceira. Cabe ressaltar que o eixo 1 retém aproximadamente 60% da inércia total (variabilidade total), sendo o principal eixo para ser analisado com o objetivo de relacionar os perfis aos pesos.

Ainda de acordo com a Tabela 8 os perfis que estão em lados opostos e que possuem a maior contribuição absoluta sobre o eixo 2 são, de um lado ,p_exc e p_pess com as contribuições absolutas de 24,7% e 20,3%, respectivamente, e, de outro lado, p_reg e p_ruim com a contribuição absoluta de 36,8% e 17,4% respectivamente. Assim pode-se perceber que os pontos extremos (maior e menor produção) estão posicionados na extremidade baixa do eixo 2 e os pontos com produção baixa para regular estão posicionados em sua extremidade alta. Assim quanto mais os perfis de linha se aproximam da extremidade alta do eixo 2, mais estão associados a um peso de touceira entre baixo e regular e quanto mais os perfis de linha se aproximam da extremidade baixa do eixo 2, mais estão associadas a valores extremos.

Para análise dos perfis de linha que representam as características físicas da cana-de-açúcar sobre o eixo 1, pode-se selecionar aqueles com a maior contribuição relativa sobre esta dimensão, conforme apresentado na Tabela 9.

Os perfis entre pobafin até poalgro e mubamed até mualmgr acrescentando-se o perfil mealfin possuem maior contribuição relativa sobre o eixo 1 e, portanto devem ser interpretados utilizando o eixo 1.

Já os perfis entre poalmgr até mubafin com exceção do mealfin possuem contribuição relativa maior sobre o eixo 2 e portanto devem ser interpretados levando-se em consideração tal eixo.

Tabela 9 - Medidas descritivas da AC sobre o perfil de linha multiplicadas por mil

| Perfil | mass (1) | qlt (2) | inr (3) | k=1 (4) | cor (5) | ctr (6) | k=2 (7) | cor (8) | ctr (9) |
|---------|-------------|------------|------------|------------|------------|------------|------------|------------|------------|
| pobafin | 59 | 869 | 139 | -1441 | 632 | 148 | -882 | 237 | 120 |
| pobamed | 102 | 971 | 153 | -1315 | 828 | 214 | -546 | 143 | 80 |
| pobagro | 67 | 869 | 66 | -1095 | 869 | 97 | 18 | 0 | 0 |
| pobamgr | 14 | 738 | 12 | -887 | 636 | 13 | 354 | 101 | 5 |
| poalfin | 6 | 890 | 6 | -1123 | 889 | 10 | -31 | 1 | 0 |
| poalmed | 35 | 959 | 30 | -1065 | 958 | 48 | -26 | 1 | 0 |
| poalgro | 45 | 656 | 45 | -779 | 440 | 33 | 546 | 216 | 35 |
| poalmgr | 21 | 947 | 11 | -362 | 173 | 3 | 764 | 774 | 32 |
| mebafin | 6 | 500 | 10 | -721 | 232 | 4 | 774 | 268 | 10 |
| mebamed | 47 | 803 | 40 | -389 | 127 | 9 | 899 | 676 | 99 |
| mebagro | 30 | 973 | 24 | -40 | 1 | 0 | 1038 | 971 | 86 |
| mebamgr | 6 | 640 | 10 | 340 | 52 | 1 | 1145 | 588 | 22 |
| mealfin | 3 | 877 | 2 | -652 | 512 | 1 | 551 | 365 | 2 |
| mealmed | 40 | 879 | 37 | 147 | 17 | 1 | 1057 | 862 | 117 |
| mealgro | 53 | 727 | 52 | 388 | 110 | 10 | 918 | 616 | 116 |
| mealmgr | 22 | 568 | 25 | 602 | 222 | 9 | 750 | 345 | 32 |
| mubafin | 18 | 703 | 4 | 301 | 292 | 2 | 357 | 411 | 6 |
| mubamed | 74 | 708 | 39 | 678 | 625 | 41 | 247 | 83 | 12 |
| mubagro | 61 | 954 | 32 | 826 | 937 | 50 | -111 | 17 | 2 |
| mubamgr | 14 | 947 | 12 | 950 | 749 | 15 | -489 | 198 | 9 |
| mualfin | 6 | 890 | 3 | 699 | 882 | 4 | -63 | 7 | 0 |
| mualmed | 81 | 959 | 54 | 878 | 829 | 76 | -347 | 130 | 26 |
| mualgro | 133 | 953 | 114 | 934 | 728 | 140 | -520 | 225 | 94 |
| mualmgr | 56 | 884 | 78 | 1026 | 541 | 71 | -817 | 343 | 97 |

(1) – massa do perfil

(2) – contribuição relativa total considerando as dimensões 1 e 2

(3) – inércia relativa

(4) – coordenada principal sobre o eixo 1

(5) – contribuição relativa sobre o eixo 1

(6) – contribuição absoluta sobre o eixo 1

(7) – coordenada principal sobre o eixo 2

(8) – contribuição relativa sobre o eixo 2

(9) – contribuição absoluta sobre o eixo 2

Analisando as contribuições relativas, já mencionadas, pode-se refazer a Figura 31 destacando aquelas que devem ser interpretadas por cada um dos eixos.

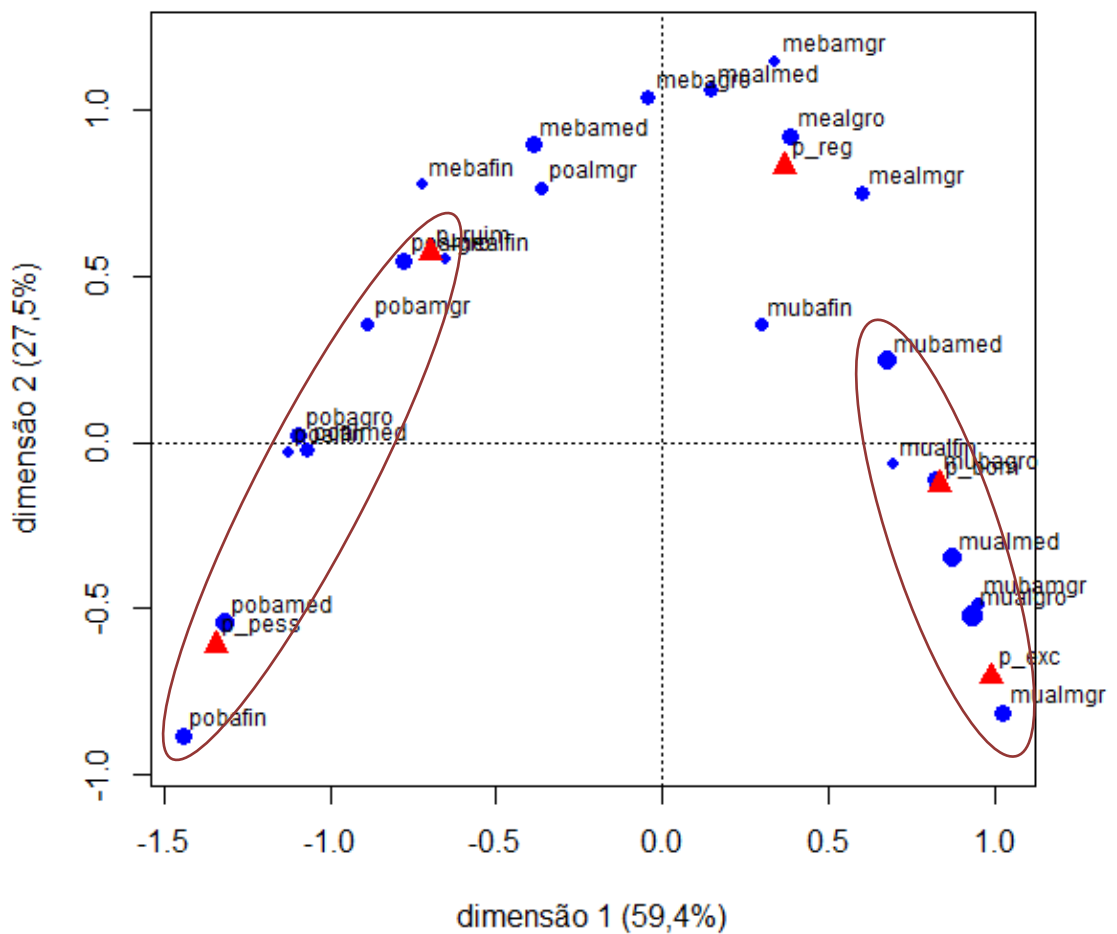


Figura 32 – Mapa perceptual simétrico das características da cana-de-açúcar e peso da touceira destacando-se os pontos que devem ser interpretados pelo eixo 1

Pode-se interpretar que os perfis compreendidos dentro da elipse a esquerda da Figura 32 podem ser interpretadas com baixa produtividade e em contrapartida os perfis compreendidos dentro da elipse a direita da Figura 32 podem ser associados a uma alta produtividade. Já os perfis que não estão compreendidos em nenhuma das elipses estão associados a uma produção regular.

Com o objetivo de relacionar perfis de linha e coluna no mesmo mapa perceptual faz-se necessário o uso do mapa assimétrico. Para relacionar cada perfil de linha (características fenotípicas) com o vértice que representa a categoria de produtividade usa-se o mapa perceptual principal por linha.

Ressalta-se que os pontos mebagro e mealmed estão mal representados no eixo 1. Tomando por base esta reordenação, percebe-se que todos os pontos até mebagro têm peso médio abaixo da média geral; os demais, acima da média geral.

Através desta ordenação e associada à proximidade dos perfis de linha aos vértices das colunas pode-se classificar os perfis de linha quanto ao peso da touceira, como mostra a Tabela 10.

Tabela 10 – Classificação dos perfis de linha quanto às faixas de peso da touceira com ordenação crescente

| Perfis abaixo da média geral | | Perfis acima da média geral | |
|------------------------------|----------------|-----------------------------|----------------|
| Peso da touceira | Característica | Peso da touceira | Característica |
| p_pess | pobafin | p_reg | mealmed |
| | pobamed | | mebamgr |
| | pobagro | | mealgro |
| p_pess e p_ruim | poalfin | ponto atípico | mealmgr |
| | poalmed | | mubafin |
| | pobamgr | | mubamed |
| p_ruim | poalgro | p_bom | mualfin |
| | mealfin | | mubagro |
| | mebafin | | mualmed |
| | poalmgr | | mubamgr |
| p_ruim e p_reg | mebamed | p_exc | mualgro |
| | mebagro | | mualmgr |

Na Figura 34, pode-se visualizar a associação do número de colmos com o peso da touceira. Assim percebe-se que o número de colmos da touceira é um fator determinante para que se possa avaliar seu peso. Touceiras com muitos colmos tendem a ter pesos maiores e touceiras com poucos colmos tendem a ter pesos menores, independentes da sua altura e diâmetro.

Essa informação é de fundamental importância, uma vez que pode facilitar os aspectos operacionais na seleção de famílias para o melhoramento genético. Isto ocorre porque o pesquisador não necessitará esperar a época da colheita para a seleção de famílias com as melhores características de peso da

peso de touceira baixo, como também as famílias de número 61 e 50 e 65 estão associadas a altos pesos de touceiras.

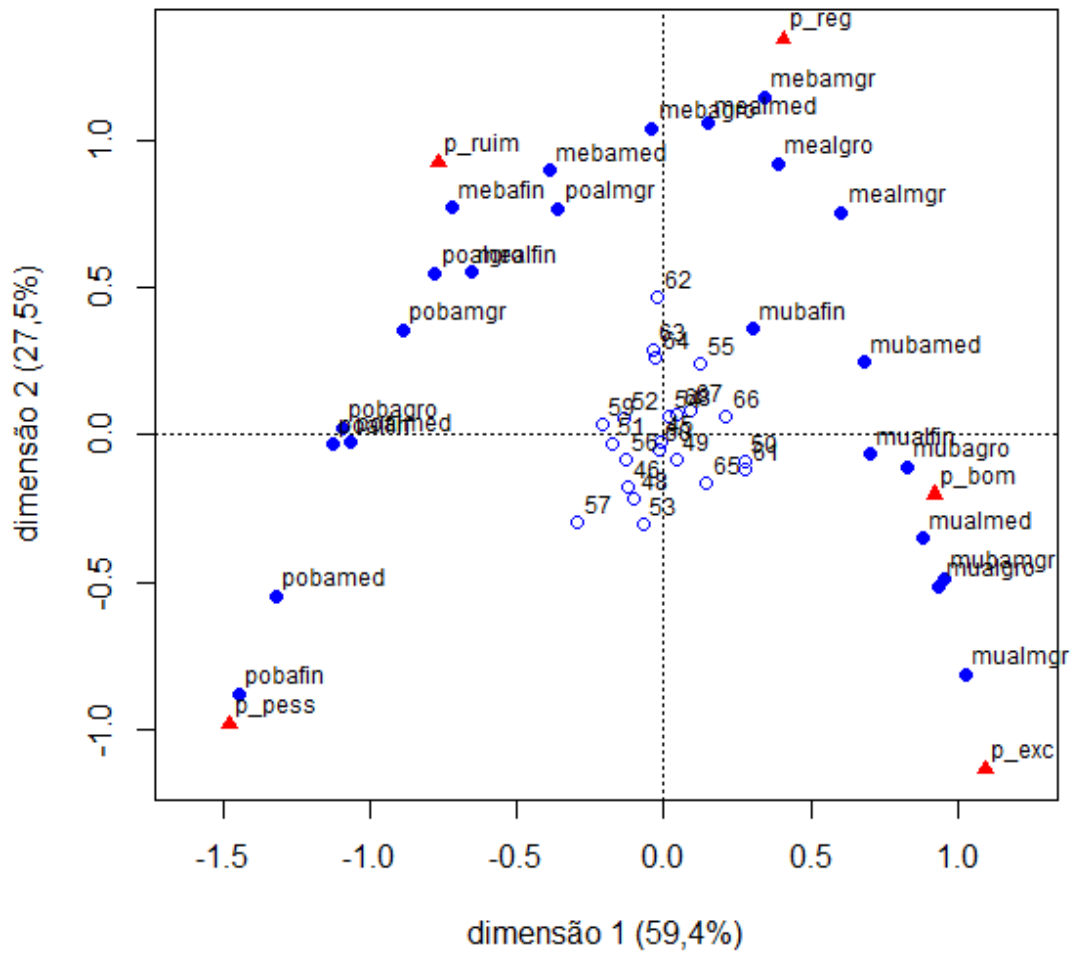


Figura 35 – Mapa perceptual principal por linha de características da cana e peso da touceira com a projeção dos pontos suplementares das famílias da cana

CONCLUSÃO

A análise de correspondência mostrou ser um método simples para análise de dados categóricos sob a forma de uma tabela de contingência, porém, sua teoria é extensa e exige um bom embasamento teórico de álgebra linear. Ler o presente trabalho exige tempo e disposição, mas ao mesmo tempo, minimiza o esforço do leitor na busca por outros trabalhos para o entendimento desta técnica. Além disso, a abordagem geométrica, associada ao exemplo ilustrativo, permite que um leitor não muito afeito aos desenvolvimentos algébricos possa ter uma visão completa do passo a passo da análise de correspondência simples.

Ao aplicar análise de correspondência aos dados da cana-de-açúcar verificou-se que o número de colmos é um fator determinante para antecipar a categoria de peso da touceira.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALCARDE, André Ricardo. **Geração de energia elétrica**. Disponível em: <www.agencia.cnptia.embrapa.br/gestor/cana-de-acucar/arvore/CONTAG01_107_22122006154841.html>. Acesso em: 15 jun. 2012.
- ALVES, L. B.; BELDERRAIN, M. C.; SCARPEL, R. A. Tratamento multivariado de dados por análise de correspondência e análise de agrupamentos. In: Encontro de Iniciação Científica e Pós Graduação do ITA – XIII ENCITA, 13, 2007, São José dos Campos. **Anais eletrônicos...** São José dos Campos: Instituto Tecnológico da Aeronáutica, 2007. Disponível em: <<http://www.bibl.ita.br/xiiiencita/MEC17.pdf>>. Acesso em: 24 jun. 2012.
- ARANHA, R. N. Análise de correspondência para avaliação do perfil de mulheres na pós-menopausa e o uso da terapia de reposição hormonal. **Cadernos de Saúde Pública**, Rio de Janeiro, vol. 20, n. 1, p. 100-108, jan-fev, 2004.
- BEH, E. J. Simple correspondence analysis: a bibliographic review. **International Statistical Review**, v. 72, n. 2, p. 257-284, ago. 2004.
- BEH, E. J. A few moments for simple correspondence analysis. In: Annual ASEARC Research Conference, 3, 2009. **Anais eletrônicos...** Newcastle: Applied Statistics Education and Research Collaboration, 2009. Disponível em: <<http://hdl.handle.net/1959.13/919141>>. Acesso em: 25 jun. 2012.
- BENZÉCRI, J.-P. **Correspondence analysis handbook**. Statistics: textbooks and monographs, v. 125. New York: Marcel Dekker, 1992.
- BERTELLI, L. G. A indústria da cana não pode parar. **Diário Comércio Indústria e Serviços**. São Paulo: 10 jan. 2012. Disponível em: <<http://www.dci.com.br/A-industria-da-cana-nao-pode-parar-4-405405.html>>. Acesso em: 15 jun 2012.
- BOLDRINI, J. L. et al. **Álgebra linear**. São Paulo: Harper & Row, 1978.
- CARVALHO, J. R. P.; VIEIRA, S. R.; MORAN, R. C. C. P. Análise de correspondência : uma ferramenta útil na interpretação de mapas de produtividade. **Revista Brasileira de Ciência do Solo**, v. 26, p. 435-443, 2002.
- CASTRO, Rogério Donizeti. **Seleção de famílias de irmãos completos via BLUPIS**. Tese (Doutorado no Programa de Pós Graduação em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa, MG, 2012.
- COMPANHIA NACIONAL DE ABASTECIMENTO – CONAB. **Cana-de-açúcar primeiro levantamento**. [online] Disponível em: <http://www.conab.gov.br/OlalaCMS/uploads/arquivos/11_12_08_11_00_54_08.pdf>. Acesso em: 15 jun. 2012.

CZERMAINSKI, ANA B.C. **Análise de correspondência**. Seminário. Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo. Piracicaba: [s.n.], 2004.

DANTAS, Carlos A. B. **Probabilidade: um curso introdutório**. 2. ed. São Paulo: EDUSP, 2004.

DE LEEUW, J.; MAIR, P. Simple and canonical correspondence analysis using the R package anacor. **Department of Statistics Papers**. Los Angeles: Department of Statistics, UCLA, 2007. Disponível em: <<http://escholarship.org/uc/item/1gf0b3m7>>. Acesso em: 25 jun. 2012.

FAVERO, L.P. et al. **Análise de dados**: modelagem multivariada para tomada de decisões. Rio de Janeiro: Elsevier, 2009.

FERREIRA, Daniel F. **Estatística multivariada**. Lavras: Editora UFLA, 2008.

FIGUEIRA, A. P. C. Faça você mesmo... procedimento ANACOR passo a passo. **Revista Iberoamericana de Educación**, v. 33, n. 8, 2004. Disponível em: <www.rieoei.org/deloslectores/795Couceiro.PDF>. Acesso em: 26 jun. 2012.

GHINI, R.; ZARONI, M. M. H. Relação entre coberturas vegetais e supressividade de solos a *Rhizoctonia solani*. *Fitopatologia Brasileira*, Brasília, vol. 26, n. 1, mar. 2001. Disponível em: <<http://dx.doi.org/10.1590/S0100-41582001000100002>>. Acesso em: 24 jun. 2012.

GREENACRE, M. The contributions of rare objects in correspondence analysis. **Barcelona GSE Working Paper**, nº 571, 2011. Disponível em: <http://research.barcelonagse.eu/One_Paper.html?paper=571>. Acesso em: 25 jun. 2012.

GREENACRE, Michael. **Correspondence analysis in practice**. *Interdisciplinary Statistics*. 2. ed. Boca Raton, FL: Chapman & Hall, 2007.

GREENACRE, Michael J. **Theory and applications of correspondence analysis**. Orlando: Academic Press, 1984.

GREENACRE, M.; BLASIUS, J. (Org.). **Multiple correspondence analysis e related methods**. *Statísticas in Social e Behavioral Science Series*. Boca Raton, FL: Chapman & Hall, 2006.

GREENACRE, M.; FARGAS, R. T. Correspondence analysis of raw data. **Ecology**, v. 91, n. 4, p. 958–963, 2010.

GREENACRE, M.; NENADIĆ, O. **Simple, multiple and joint correspondence analysis**. Version 0.33, 2010. Disponível em: <<http://www.carme-n.org/>>.

GUEDES, T. A. et al. Seleção de variáveis categóricas utilizando análise de correspondência e análise procrustes. **Acta Scientiarum**, Maringá, v. 21, n. 4, p. 861-868, 1999.

GUEDES, T. A.; IVANQUI, I. L. Aspecto da seleção de variáveis na análise de correspondência. In: Conferência Internacional "Experiências e Expectativas do Ensino de Estatística - Desafios para o Século XXI", 1999. **Atas...** Florianópolis, 1999. Disponível em: <<http://www.inf.ufsc.br/cee/posters/ResumoFloripa.html>>. Acesso em: 25 jun. 2012.

HILDEBRAND, K.F.; MÜLLER-FUNK, U. HOMALS for dimension reduction in information retrieval. In: GAUL, W. et al. **Challenges at the interface of data analysis, computer, science and optimization**. New York: Springer, 2012.

JELIHOVSCHI, E, G.; FERRAZ, M. I. Análise do conjunto dos candidatos ao vestibular da UESC no ano de 2008 usando análise de correspondência. **Revista Brasileira de Biometria**, São Paulo, v. 28, n. 4, p. 117-136, 2010.

KHATTREE, R.; NAIK, D. N. **Multivariate data reduction and discrimination: with SAS[®] software**. Books By Users. Cary, NC: SAS Institute Inc.,2000.

LAY, David C. **Álgebra linear e suas aplicações**. 2. ed. Rio de Janeiro: LTC, 1999.

LEON, Steven J. **Álgebra linear com aplicações**. 4. ed. Rio de Janeiro: LTC, 1999.

LOPES, L. G. Águas subterrâneas de região produtora de região produtora de cana-de-açúcar paulista: relação entre indicadores de qualidade e características do poço. **Nucleus**, Ituverava, v. 7, n. 1, 2010. Disponível em: <<http://www.nucleus.feituverava.com.br/index.php/nucleus/article/view/327>>. Acesso em: 26 jun. 2012.

MATOS, I. M. T.; AMARAL, J. Aula 33. In: Matos, I.M.T. **Tópicos de álgebra linear**. [Apostila]. Departamento de Engenharia de Eletrônica e Telecomunicações e de Computadores, Instituto Superior de Engenharia de Lisboa. Disponível em: <http://www.deetc.isel.ipl.pt/paginaspeessoais/isabelteixeira/MatAp/AL_ApontamentosTeoria.pdf>. Acesso em: 23 mar. 2012.

MURDOCH, David, C. **Geometria Analítica**: com uma introdução ao cálculo vetorial e matrizes. 2.ed. Rio de Janeiro: LTC, 1975.

NAITO, Sérgio D. N. P. **Análise de correspondência generalizada**. Dissertação (Mestrado em Bioestatística) – Universidade de Lisboa. Lisboa, 2007.

NYFJÄLL, M. **Aspects on correspondence analysis plots under complex survey sampling designs**. Research Report, Department of Information Science, division of Statistics, Uppsala University, n. 2, 2002a.

NYFJÄLL, M. **On correspondence analysis under complex sampling designs**. Thesis (Degree of Licentiate of Philosophy (Fil. Lic.) in Statistic) – Uppsala University. Uppsala: 2002b.

PAMPLONA, Admur S. **Análise de correspondência para dados com estrutura de grupo**. Dissertação (Mestrado em Estatística) – Instituto de

Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas. Campinas, 1998.

PETERNELLI, L. A.; MELLO, M. P. **Conhecendo o R: uma visão estatística**. Série didática. Viçosa, MG: Editora UFV, 2011.

POOLE, David. **Álgebra linear**. São Paulo: Pioneira Thomson Learning, 2004.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. R Foundation for Statistical Computing. Vienna, Austria: 2011. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org/>>.

REGAZZI, Adair J., **EST 640 – Modelos lineares I**. [Apostila]. Departamento de Estatística, Universidade Federal de Viçosa. Viçosa, MG: [s.n.], 2010.

SILVA, J. C. G. L. et al. O uso da análise de correspondência e de *cluster* para a percepção das relações no comércio internacional: o caso do setor de móveis sul-brasileiro e as barreiras à Alca. **Revista de Administração da Universidade de São Paulo – RAUSP**, São Paulo, v. 43, n. 1, p. 44-58, jan./fev./mar. 2008.

SILVA NETO, P. V.; RAPOSO, M. C. F. Análise de agrupamentos usada como auxiliar na interpretação dos resultados da análise de correspondência múltipla: um exemplo prático. In: Simpósio Nacional de Probabilidade e Estatística - Sinape, 19, 2010, São Pedro. **Anais eletrônicos...** Campinas: Universidade Estadual de Campinas, 2010. Disponível em: <<http://www.ime.unicamp.br/sinape/19sinape/node/506>>. Acesso em: 24 jun. 2012.

SOUZA, Augusto C. **Análise de correspondência aplicada à ECINF: a diversidade do setor informal urbano no Brasil**. Dissertação (Mestrado em Estudos Populacionais e Pesquisas Sociais) – Escola Nacional de Ciências Estatísticas. Rio de Janeiro, 2004.

SOUZA, A. C.; BASTOS, R. R.; VIEIRA, M. T. Análise de correspondência simples e múltipla para dados amostrais complexos. In: Simpósio Nacional de Probabilidade e Estatística - Sinape, 20, 2011, São Pedro. **Anais eletrônicos...** Campinas: Universidade Estadual de Campinas, 2011. Disponível em: <<http://www.ime.unicamp.br/sinape/sites/default/files/Artigo%20Sinape%20v2.pdf>>. Acesso em: 24 jun. 2012.

SOUZA, M. L. M.; BASTOS, R. R.; VIEIRA, M. T. Escores ponderados a partir da solução HOMALS. In: Simpósio Nacional de Probabilidade e Estatística - Sinape, 19, 2010, São Pedro. **Anais eletrônicos...** Campinas: Universidade Estadual de Campinas, 2010. Disponível em: <http://www.ime.unicamp.br/sinape/sites/default/files/Souza_Bastos_Vieira_v7.pdf>. Acesso em: 24 jun. 2012.

SOUZA, V. B. C; PETERNELLI, L. A., BERNARDES, D. P. Seleção de famílias de cana-de-açúcar com base na categorização de variáveis. **Resumo estendido**. 2012.