

XVIII Jornadas de Classificação e Análise de Dados
Universidade de Trás-os-Montes e Alto Douro, 7 a 9 de Abril de 2011

Novas Aplicações de Métodos Multivariados na Análise da Mortalidade: um Estudo na Região Norte de Portugal, 2001-2005

Lara Teixeira¹, Vasco Machado², Manuela Felício³, Arminda Manuela Gonçalves⁴

¹Departamento de Matemática e Aplicações – Universidade do Minho

Campus de Azurém, 4800-058 Guimarães, lara-teixeira@hotmail.com ;

²Departamento de Saúde Pública – Administração Regional de Saúde do Norte, I.P.

Rua Anselmo Braamcamp, 144, 4000-078 Porto, vmachado@arsnorte.min-saude.pt ;

³Departamento de Saúde Pública – Administração Regional de Saúde do Norte, I.P.

Rua Anselmo Braamcamp, 144, 4000-078 Porto, mfelicio@arsnorte.min-saude.pt ;

⁴Departamento de Matemática e Aplicações – Universidade do Minho

Campus de Azurém, 4800-058 Guimarães, mneves@math.uminho.pt

Sumário

A análise da mortalidade é fundamental no processo de planeamento da saúde e dos serviços de saúde, sendo a taxa de mortalidade padronizada pela idade um dos principais indicadores utilizados. A aplicação da Análise em Componentes Principais e da Análise de *Clusters* surge, neste trabalho, com o objectivo de identificar conjuntos de causas de morte que possam estar mais correlacionadas entre si e de reunir Agrupamentos de Centros de Saúde (ACES) segundo perfis de mortalidade semelhantes.

Palavras-chave Agrupamento de Centros de Saúde (ACES), análise de *Clusters*, análise em componentes principais, taxa de mortalidade padronizada

1. Introdução

A mortalidade é considerada uma medida directa das necessidades em cuidados de saúde, reflectindo a carga global da doença na população, não só em termos da incidência da doença, como da capacidade de a tratar. Daí a importância dos indicadores de mortalidade no processo de planeamento da saúde e dos serviços de saúde.

A taxa de mortalidade padronizada pela idade é um dos indicadores usados na análise da mortalidade. A sua utilização, para um conjunto de causas de morte específicas, permite a comparação dos seus valores entre diferentes unidades territoriais. No entanto, os problemas de saúde não são balizados por fronteiras geográficas e, muito menos, administrativas. Um dos objectivos deste estudo é argumentar neste sentido. A Análise em Componentes Principais e a Análise de *Clusters*, métodos da Análise de Dados Multivariados, são usadas neste estudo com o objectivo de identificar conjuntos de causas de morte que possam estar mais correlacionadas

entre si e, também, conjuntos de Agrupamentos de Centros de Saúde (ACES) da região Norte (respectivas áreas geodemográficas), cujo perfil de mortalidade seja mais semelhante.

2. Base de dados e metodologia

É utilizada a base de dados regional dos óbitos para calcular a taxa de mortalidade padronizada (TMP) pela idade de um conjunto de 13 causas de morte específicas, para 24 ACES da região Norte. Calcula-se a TMP média anual no quinquénio 2001-2005 apresentando-se aqui apenas as TMP calculadas para ambos os sexos. Obtém-se uma matriz $X = (x_{ij})$, de dimensão 24×13 , onde x_{ij} representa o valor da TMP na variável de ordem j (causa de morte específica) observada no indivíduo i (ACES).

2.1. Taxa de mortalidade padronizada pela idade (TMP)

O método directo de padronização consiste na aplicação das taxas específicas de mortalidade por idade a uma população padrão cuja composição etária é fixa, distribuindo-se pelos mesmos grupos etários das taxas específicas. O método consiste, portanto, em calcular as taxas de mortalidade esperadas na população padrão. De acordo com um procedimento corrente na literatura escolheu-se a população padrão europeia, com grupos etários decenais. Para mais considerações ver Dobson & al. (1991) e Fay & Feur (1997).

Utiliza-se um teste de hipóteses para a comparação das taxas de mortalidade padronizadas de duas populações para verificar se a diferença nos valores da TMP entre duas áreas geográficas é estatisticamente significativa. Obtém-se uma classificação dos ACES, para cada causa de morte, em quatro classes: a TMP de um dado ACES é inferior ou superior à TMP da região, com ou sem significância estatística. São produzidos mapas da região que permitem a visualização e identificação das diferenças testadas.

2.2. Análise em componentes principais

A Análise em Componentes Principais (ACP) tem por objectivo encontrar um conjunto de novas variáveis não correlacionadas linearmente, denominadas componentes principais, em que cada uma destas componentes pode ser escrita como uma combinação linear das variáveis originais (neste caso, as causas de morte específicas) (Reis (2001)).

O *biplot* (Gabriel (1971)) é uma representação gráfica de dados multivariados que permite a sobreposição dos indivíduos e das variáveis. A sobreposição das projecções das duas nuvens no mesmo plano torna mais fácil a interpretação, desde que se compreenda que as nuvens têm significados diferentes.

2.3. Análise de *Clusters*

A Análise de *Clusters* tem como principal objectivo agrupar um conjunto de objectos a classificar (variáveis ou indivíduos) num número pequeno de classes, que reflectam as relações de semelhança e/ou oposição entre esses objectos (Branco (2004)). Aplica-se o método hierárquico, com recurso ao coeficiente de correlação de *Pearson* para a comparação de variáveis (causas de morte) e à distância Euclidiana para a comparação de pares de indivíduos (ACES), utilizando vários critérios de agregação. Assim, procura-se determinar grupos homogêneos de ACES da região Norte considerando todas as causas e, também, classes de causas de morte entretanto agrupadas. Uma vez constituídos os *clusters* de ACES, identificam-se os que diferem relativamente às variáveis que contribuem para o seu agrupamento, utilizando-se os testes não-paramétricos de *Kruskal-Wallis* e de *Tukey* (Higgins (2004)).

3. Resultados

O cálculo da taxa de mortalidade padronizada (TMP) pela idade e o teste de hipóteses aplicado permitiu, para uma determinada causa de morte específica, identificar os ACES que apresentam valores da TMP significativamente diferentes do valor da TMP da região. Permitiu, também, para um determinado ACES da região Norte, identificar quais as causas de morte em que esse ACES apresenta valores da TMP significativamente diferentes dos valores médios da região (ARS Norte (2008)).

Com a aplicação da ACP pode concluir-se quais as variáveis (causas de morte) que estão mais correlacionadas com cada uma das componentes principais retidas e quais os indivíduos (ACES) que mais contribuíram para a formação dessas componentes. Foram retidas 4 componentes que explicam 72,3% da variância total. Na Figura 1 pode observar-se o *biplot* da representação das variáveis e indivíduos no plano composto pela primeira e segunda componentes principais.

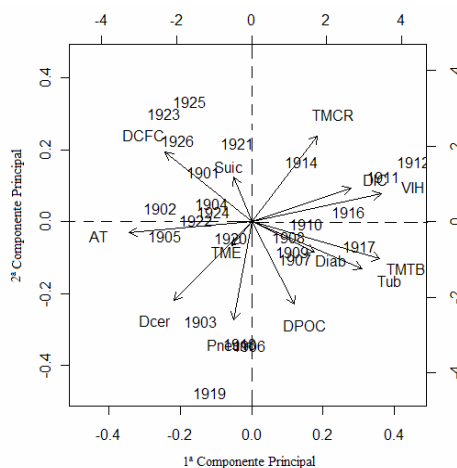


Figura 1: *biplot* da 1ª e 2ª componentes.

Como exemplo de resultados encontrados refira-se que a aplicação da ACP e da análise de *Clusters* permitiu identificar os ACES Matosinhos, Porto Ocidental e Oriental, Gondomar e Valongo como sendo os que apresentam valores da TMP mais elevados nas causas de morte tuberculose, VIH/sida, tumor maligno da traqueia, brônquios e pulmão e doença isquémica do coração.

4. Conclusão

Os procedimentos multivariados adoptados neste estudo foram muito úteis para identificar padrões homogéneos no comportamento das causas de mortalidade e dos ACES. Permitiram reduzir a dimensão do conjunto de informação e agrupar ACES segundo perfis de mortalidade semelhantes. Espera-se que os resultados deste trabalho possam apoiar um planeamento em saúde e das intervenções em saúde integrado e mais balizado pelas *fronteiras epidemiológicas* do que pelas fronteiras geográficas e/ou administrativas.

Referências:

Livros:

BRANCO, J.A. (2004) *Uma Introdução à Análise de Clusters*, Évora, Sociedade Portuguesa de Estatística.

HIGGINS, J.J. (2004). *Introduction to Modern Nonparametric Statistics*. Duxbury Advanced Series.

REIS, E. (2001). *Estatística Multivariada Aplicada*. Lisboa, Edições Sílabo, 2ª edição.

Artigos:

DOBSON, A. & al. (1991). Confidence intervals for weighted sums of Poisson parameters. *Statistic in Medicine*, 10, 457-462.

GABRIEL, K.R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, vol.58, n.03, 453–467.

FAY, M.P. & FEUER, E.J. (1997) Confidence intervals for directly adjusted rates: a method based on the gamma distribution. *Statistic in Medicine*, 16, 791-801.

Endereços de Web:

ARS Norte (2008). mort@lidades.geres – Mortalidade Geral e Específica, Região Norte 2001-2005, http://portal.arsnorte.min-saude.pt/ARSNorte/dsp/AM_0105, (acedido em 1 de Janeiro 2011).