

ME731 - Métodos em Análise Multivariada

– Análise de Correspondência II –

Prof. Carlos Trucíos
ctrucios@unicamp.br
ctruciosm.github.io

Instituto de Matemática, Estatística e Computação Científica,
Universidade Estadual de Campinas

Aula 18



Agenda I

- 1 Introdução
- 2 Análise de Correspondência Múltipla

Introdução

Introdução

- AC é uma técnica para representar as categorias de duas variáveis qualitativas em (usualmente) um gráfico bidimensional.

Introdução

- AC é uma técnica para representar as categorias de duas variáveis qualitativas em (usualmente) um gráfico bidimensional.
- O que acontece se estivermos interessados em mais do que duas variáveis qualitativas?

Introdução

- AC é uma técnica para representar as categorias de duas variáveis qualitativas em (usualmente) um gráfico bidimensional.
- O que acontece se estivermos interessados em mais do que duas variáveis qualitativas? Não podemos mais utilizar AC.

Introdução

- AC é uma técnica para representar as categorias de duas variáveis qualitativas em (usualmente) um gráfico bidimensional.
- O que acontece se estivermos interessados em mais do que duas variáveis qualitativas? Não podemos mais utilizar AC.
- Análise de Correspondência Múltipla (ACM) é uma generalização da AC e estende este último para três ou mais variáveis.

Análise de Correspondência Múltipla

Análise de Correspondência Múltipla

ACM

Através da ACM obtemos um gráfico (geralmente bi-dimensional) com a informação contida em uma Tabela de contingência multidimensional, de forma que preservemos, o máximo possível, a relação existente entre as categorias das variáveis.

Análise de Correspondência Múltipla

ACM

Através da ACM obtemos um gráfico (geralmente bi-dimensional) com a informação contida em uma Tabela de contingência multidimensional, de forma que preservemos, o máximo possível, a relação existente entre as categorias das variáveis.

Suponha n observações das quais temos informação referente a p variáveis categóricas X_1, \dots, X_p com k_1, \dots, k_p categorias cada uma.

Análise de Correspondência Múltipla

ACM

Através da ACM obtemos um gráfico (geralmente bi-dimensional) com a informação contida em uma Tabela de contingência multidimensional, de forma que preservemos, o máximo possível, a relação existente entre as categorias das variáveis.

Suponha n observações das quais temos informação referente a p variáveis categóricas X_1, \dots, X_p com k_1, \dots, k_p categorias cada uma.

Definimos a matriz indicadora $\mathbf{G}_{n \times (k_1 + \dots + k_p)}$. Os elementos de \mathbf{G} são zeros e uns (uns indicam que a observação pertence a determinada categoria).

Análise de Correspondência Múltipla

ACM

Através da ACM obtemos um gráfico (geralmente bi-dimensional) com a informação contida em uma Tabela de contingência multidimensional, de forma que preservemos, o máximo possível, a relação existente entre as categorias das variáveis.

Suponha n observações das quais temos informação referente a p variáveis categóricas X_1, \dots, X_p com k_1, \dots, k_p categorias cada uma.

Definimos a matriz indicadora $\mathbf{G}_{n \times (k_1 + \dots + k_p)}$. Os elementos de \mathbf{G} são zeros e uns (uns indicam que a observação pertence a determinada categoria).

ACM implica aplicar AC na matriz \mathbf{G} .

Análise de Correspondência Múltipla

- Aplicar AC na matriz \mathbf{G} é equivalente a aplicar AC na matriz $\mathbf{G}'\mathbf{G}$ (lembre-se AC aplica SVD em $\mathbf{G} = \mathbf{U} \mathbf{D} \mathbf{V}'$ e a decomposição espectral em $\mathbf{G}'\mathbf{G}$ nos dará o mesmo \mathbf{V}).
- A matriz $\mathbf{G}'\mathbf{G}$ é chamada de matriz de Burt.
- A matriz de Burt ($\mathbf{G}'\mathbf{G}$) tem blocos quadrados na diagonal para cada variável e tem, fora da diagonal, blocos retangulares para cada par de variáveis.
- Os blocos quadrados na diagonal são as frequências de cada categoria para a respectiva variável.
- Os blocos retangulares fora da diagonal são as tabelas de contingência para cada par de variáveis.

Análise de Correspondência Múltipla

Interpretação

- Se dois pontos representando categorias da mesma variável estão perto, essas categorias podem ser combinadas em uma só.

Referências

Referências

- Härdle, W. K., & Simar, L. (2019). Applied Multivariate Statistical Analysis. Fifth Edition. Springer Nature. Capítulo 15.
- Jhonson, R. & Wichern, D. (2007). Applied Multivariate Statistical Analysis. Sixth Edition. Person. Capítulo 12.7
- Mingoti, S. (2007). Análise de Dados Através de Métodos de Estatística Multivariada: Uma abordagem aplicada. Editora UFMG. Capítulo 8.