

Tópicos Avançados em Regressão

Plano de Ensino

Gilvan R. Guedes

Departamento de Demografia - UFMG
Sala: 3093 - Ramal: 7165
Atendimento e consulta: sob agendamento
Email: grguedes@cedeplar.ufmg.br

16 de outubro de 2013

1 Aulas

Horário: 09:30h as 12:30h

Local: Sala 3138 - FACE/UFMG

Datas: Outubro (18, 25); Novembro (01, 08, 22, 29); Dezembro (06, 13).

2 Objetivo

Este curso tem caráter predominantemente aplicado (empírico). O objetivo é fornecer ao aluno os conhecimentos teóricos e operacionais básicos sobre modelos estatísticos multivariados e de regressão comumente aplicados a análises de fenômenos demográficos e econômicos. Literatura específica aplicada às diferentes metodologias devem ser estudadas, assim como a participação nas práticas laboratoriais para que o objetivo seja plenamente cumprido. Ao final do curso, espera-se que o aluno tenha a capacidade de manipular dados nos programas relativos a cada técnica (Stata, R, Latent Gold, GoM 3.4, ugom) e que saiba aplicar esses métodos a fenômenos de interesse na área de Demografia, Economia e Saúde.

3 Avaliação

Os 100 pontos serão distribuídos como a seguir:

- 40 pontos para aplicação empírica do método GoM
- 60 pontos para proposta de aplicação empírica dos métodos aplicados à área de interesse do aluno (máximo de 10 páginas)

Data de entrega: a combinar.

4 Grade of Membership

O curso começará com a apresentação e prática do método Grade of Membership. Esta seção será dividida em três subseções, compreendendo 2 aulas teóricas e 1 aula de laboratório. Esta parte será ministrada pelos Professores Gilvan Guedes (Departamento de Demografia) e Mário Rodarte (Departamento de Economia).

4.1 Centróide, Lógica Nebulosa e Partição Difusa

Nesta parte será introduzida a ideia de centróide e uma introdução a noção de lógica fuzzy. A lógica fuzzy será abordada com exemplos de partição difusa em contextos de conjuntos matemáticos com funções de pertinência não-binárias.

Bibliografia sugerida: (Giles, 1988); (Manton et al., 1994); (Correia et al., 1996); (Guedes et al., 2014)

4.2 O Método Grade of Membership

Nesta parte será apresentada uma introdução ao Método Grade of Membership em sua versão para criação de perfis multidimensionais aplicados a dados de período (*cross-section*).

Serão discutidas as propriedades dos estimadores de Máxima Verossimilhança, as condições para obtenção de ótimos globais, e as fórmulas e os testes para se encontrar o melhor ajuste aos dados. Serão também discutidas as formas de criação de tipos mistos, utilizando os estimadores de partição difusa.

Bibliografia sugerida:

- *Modelo Teórico:* (Manton et al., 1994); (Guedes et al., 2014)
- *Propriedades dos Estimadores:* (Woodbury and Clive, 1974); (Caetano and Machado, 2009); (Guedes et al., 2010); (Guedes et al., 2011); (Guedes et al., 2011)
- *Fórmulas e testes de ajuste:* (Akaike, 1992); (Manton et al., 1994)
- *Algoritmos para tipos mistos:* (Woodbury et al., 1978); (Sawyer et al., 2002); (Guedes et al., 2014)

4.3 Aplicações Práticas do GoM

Nesta parte são apresentados vários artigos publicados, utilizando o Método GoM em diferentes áreas, como migração, saúde, mortalidade, sistemas de uso do solo, pobreza multidimensional, e mercado de trabalho. Os estudos vão de teses, dissertações e textos para discussão a artigos científicos nacionais e internacionais.

Apesar de já há alguns anos sendo utilizada para analisar empiricamente base de dados de alta dimensionalidade, novas propostas vem sendo incorporadas à literatura empírica, revelando formas criativas e flexíveis para a utilização do método. Exemplos dessas novas formas de utilização do GoM:

- redução de dimensionalidade,
- pesquisas quanti-quali,
- indicadores de hierarquias,
- pobreza e desigualdade multidimensionais,
- marcadores multidimensionais de saúde.

Bibliografia sugerida:

- *Aplicações empíricas:* (Garcia et al., 2007); (Melo, 2007); (Assis et al., 2008); (Alves et al., 2008); (Andreotti et al., 2009); (Guedes et al., 2009)
- *Redução de dimensionalidade:* (Guimarães et al., 2010); (Guedes and Carmo, 2012)
- *Pesquisas quanti-quali:* (Miranda-Ribeiro et al., 2007); (Guedes, 2010)
- *Indicadores de hierarquia:* (Guedes et al., 2009); (Reis, 2009)
- *Indicadores multidimensionais de bem-estar:* (Guedes et al., 2012); (Barbieri et al., 2013); (Hull and Guedes, 2013)
- *Marcadores de saúde:* (Baptista, 2003); (Guedes et al., 2013)

4.4 Prática de Laboratório

O aluno terá a oportunidade de trabalhar com uma base de dados demográfica (Censos históricos) para explorar as potencialidades do método Grade of Membership. Utilizando os recursos computacionais do Laboratório de Computação Aplicada (LCA) do Cedeplar/UFMG, cada aluno deverá gerar um modelo GoM utilizando os dados fornecidos. Dois professores, Gilvan Guedes e Mário Marcos Rodarte, acompanharão a aula para tirar dúvidas operacionais.

Bibliografia sugerida: (Guedes et al., 2014)

5 Modelos de Classe Latente

Nesta seção o aluno deverá aprender a trabalhar com modelos com funções de pertinência nebulosa, caracterizados por agrupamento probabilístico. Esses modelos assumem que a variável latente é categórica e formada por uma série de indicadores observáveis. Para esse curso serão vistos apenas os modelos latentes de agrupamento sem covariáveis, além dos modelos de agrupamento com covariáveis ativas e inativas. Esta seção será dividida em três subseções, compreendendo 2 aulas teóricas e 1 aula de laboratório.

5.1 Motivando a Abordagem de Classe Latente

Nesta parte serão contemplados os seguintes pontos:

- Ideias básicas de análise de classe latente
- O modelo geral de probabilidade para variáveis categóricas
- Determinação do número de classes ou grupos
- Medidas de ajuste, especificação do modelo e estratégias de seleção
- Classificação de casos em segmentos de

classe latente

- Interpretação das saídas do programa Latent GOLD
- Exemplo de análise de levantamento amostral

Bibliografia sugerida:

- *Literatura técnica:*(Magidson and Vermunt, 2001); (Vermunt and Magidson, 2001); (Magidson and Vermunt, 2002); (Vermunt and Magidson, 2005a); (Vermunt and Magidson, 2005b)
- *Literatura aplicada:* (Macmillan and Eliason, 2003); (Moisio, 2004); (Macmillan and Copher, 2005); (Reis and Guedes, 2012)

5.2 Modelos de Agrupamentos Latentes - Com e Sem Covariáveis

Nesta parte o aluno deverá aprender a gerar um modelo de agrupamento probabilístico em que covariáveis são introduzidas. Essas covariáveis podem ou não afetar a probabilidade de pertencimento a uma classe, sendo por isso chamadas de ativas - quando influenciam o *class membership* e inativas - quando são utilizadas apenas como variáveis de estratificação.

Bibliografia sugerida: (Vermunt and Magidson, 2005a); (Vermunt and Magidson, 2005b)

5.3 Prática de Laboratório

O aluno terá a oportunidade de trabalhar com uma base de dados primária (sob domínio do Prof. Gilvan Guedes) para explorar as potencialidades do método de Classe Latente. Utilizando os recursos computacionais do Laboratório de Computação Aplicada (LCA) do Cedeplar/UFMG, os alunos deverão trabalhar em grupo para tentar ajustar modelos de classe latente para expressar a atitude ambiental dos indivíduos no banco de dados. Os resultados deverão ser interpretados e debatidos em aula.

Bibliografia sugerida: (Vermunt and Magidson, 2005a); (Vermunt and Magidson, 2005b); (Guedes et al., 2012)

6 Modelos de Seleção Amostral

Os modelos de seleção amostral são utilizados toda vez que estamos diante de amostras não-aleatórias em que ocorra uma seleção amostral endógena. É necessário, neste caso, um procedimento probabilístico para controlar a probabilidade de ser selecionado para a amostra observada (não-aleatória). Esse procedimento é conhecido como procedimento de correção de Heckman, o qual representa uma estimação dos estimadores em dois estágios: a probabilidade predita de seleção (primeiro estágio) e os parâmetros de interesse sobre a variável independente com seletividade endógena, corrigida pelo viés de seleção (no segundo estágio). Esta seção será dividida em 2 partes, 1 correspondendo à apresentação teórica dos modelos de seleção amostral, e 1 relativa à prática de laboratório para execução dos modelos de seleção aplicadas a microdados.

6.1 Amostras Não-aleatórias e Seleção Amostral

Nesta seção serão tratados os problemas estatísticos gerados por amostras não-aleatórias, representadas pela seleção amostral. Os dados faltantes (missing), por exemplo, são problemáticos quando resultam de uma amostra não aleatória da população. Isso ocorre porque a amostra não é uma representação aleatória da população, causando violação de um pressuposto básico do modelo de regressão linear – a amostragem aleatória – com consequências sobre a estimação MQO. Aprenderemos quando a amostra não-aleatória gera seleção amostral exógena e endógena e como solucionar os

problemas estatísticos causados pela seleção amostral endógena.

Bibliografia sugerida:

- *Bibliografia básica:* (Heckman, 1979); (Manddala, 1983); (Greene, 2000); (Wooldridge, 2002); (Cameron, 2005)

- *Bibliografia complementar:* (Hausman and Wise, 1977); (Heckman and Smith, 1995); (Heckman et al., 1997); (Heckman et al., 1997); (Long, 1997)

6.2 Prática de Laboratório

O aluno terá a oportunidade de trabalhar com uma base de dados secundária (PNAD-2011) para explorar as potencialidades do método de seleção amostral de Heckman. Utilizando os recursos computacionais do Laboratório de Computação Aplicada (LCA) do Cedeplar/UFMG, os alunos deverão trabalhar em grupo para tentar ajustar um modelo que estime o efeito da educação e da experiência sobre os salários dos trabalhadores ocupados no Brasil. Os resultados deverão ser interpretados e debatidos em aula.

Referências

- Akaike, H. (1992). Information theory and an extension of the maximum likelihood principle. *Breakthroughs in statistics*, 610–624.
- Alves, L. C., I. C. Leite, and C. J. Machado (2008). Perfis de saúde dos idosos no brasil: análise da pesquisa nacional por amostra de domicílios de 2003 utilizando o método *grade of membership*. *Cad. saúde pública* 24(3), 535–546.
- Andreotti, A., N. Minicuci, P. Kowal, and S. Chatterji (2009). Multidimensional profiles of health status: An application of the

- grade of membership model to the world health survey. *PLoS one* 4(2), e4426.
- Assis, H. M., C. J. Machado, and R. N. Rodrigues (2008). Perfis de mortalidade neonatal precoce: um estudo para uma maternidade pública de belo horizonte (mg), 2001-2006. *Rev. bras. epidemiol* 11(4), 675–686.
- Baptista, D. B. D. A. (2003). Idosos no município de são paulo: expectativa de vida ativa e perfis multidimensionais de incapacidade a partir da sabe. Dissertação (mestrado em demografia), Universidade Federal de Minas Gerais.
- Barbieri, A. F., E. Domingues, K. V. M. S. Noronha, B. L. Queiroz, G. R. Guedes, J. I. R. Rigotti, F. L. C. Feres, G. P. Motta, F. M. Cortezzi, U. Confalonieri, and K. Souza (2013). Population transitions and climate change: a multidimensional index of population vulnerability in brazil. In *Proceedings of the IUSSP Seminar on Population Dynamics and Human Dimensions of Climate Change*, Canberra, Australia. IUSSP.
- Caetano, A. J. and C. J. Machado (2009). Consistência e identificabilidade no modelo grade of membership: uma nota metodológica. *Revista Brasileira de Estudos de População* 26(1), 145–149.
- Cameron, A. C. (2005). *Microeconomics: methods and applications*. Capítulo 16. Cambridge, England: Cambridge university press.
- Correia, J. D., N. C. C. Silva, L. C. Oliveira, and L. F. Ferreira (1996). Centro geodésico e centróides: uma abordagem conceitual. *Cartografia e Cadastro* 5, 1–5.
- Garcia, R. A., B. S. Soares-Filho, and D. O. Sawyer (2007). Socioeconomic dimensions, migration, and deforestation: An integrated model of territorial organization for the brazilian amazon. *Ecological Indicators* 7(3), 719–730.
- Giles, R. (1988). The concept of grade of membership. *Fuzzy Sets and Systems* 25(3), 297–323.
- Greene, W. H. (2000). *Econometric Analysis*. Capítulo 20. (4th ed.). New Jersey: Prentice Hall.
- Guedes, G. R. (2010). Ciclo de vida e mudança na cobertura e no uso do solo na amazônia brasileira o caso da área de estudo de altamira. Tese (doutorado em demografia), Universidade Federal de Minas Gerais.
- Guedes, G. R., E. S. Brondizio, A. F. Barbieri, A. C. Resende, R. Penna-Firme, and A. O. D’Antona (2012). Poverty and inequality in the rural brazilian amazon: a multidimensional approach. *Human ecology* 40(1), 41–57.
- Guedes, G. R., A. J. Caetano, C. J. Machado, and E. S. Brondizio (2010). Identificabilidade e estabilidade dos parâmetros no método grade of membership (gom): considerações metodológicas e práticas. *Revista Brasileira de Estudos de População* 27(1), 21–33.
- Guedes, G. R. and R. L. Carmo (2012). *Urbanization and the Global Environment*, Chapter Socioeconomic and Residential Differences in Environmental Perception and Behavior: Insights from Metropolitan Brazil. New York: Nova Publishers.
- Guedes, G. R., S. M. Costa, and E. S. Brondizio (2009). Revisiting the hierarchy of urban areas in the brazilian amazon: a multilevel approach. *Population and environment* 30(4-5), 159–192.
- Guedes, G. R., C. Guimaraes, and L. P. Terra (2013). Positive attitude and wellbeing: A life cycle analysis of individuals healthy and

- happy life expectancy for brazil and mexico. In *Proceedings of the XXVII IUSSP International Population Conference*, Busan, Republic of Korea. IUSSP.
- Guedes, G. R., A. M. Hora, C. A. Dias, E. O. Braga, and M. M. Soares (2012). Avaliação crítica da Escala do Novo Paradigma Ecológico em contextos de baixo desenvolvimento. In Guedes, G. R. and Ojima, R. (Eds.), *Território, Mobilidade Populacional e Ambiente*. Governador Valadares: Editora Univale.
- Guedes, G. R., B. L. Queiroz, and L. K. VanWey (2009). Transferências intergeracionais privadas na amazônia rural brasileira. *Nova Economia* 19(2), 325–357.
- Guedes, G. R., P. C. L. Siviero, A. J. Caetano, C. J. Machado, and E. S. Brondizio (2011). Incorporando a variabilidade no processo de identificação do modelo de máximo global no grade of membership (gom): considerações metodológicas. *Revista Brasileira de Estudos de População* 28(2), 337–347.
- Guedes, G. R., P. C. L. Siviero, and C. J. Machado (2011). Aspectos práticos na identificação de um modelo grade of membership (gom) de máximo global: o uso da moda das probabilidades estimadas. *Revista Brasileira de Estudos de População* 28(2), 473–478.
- Guedes, G. R., P. C. L. Siviero, C. J. Machado, and M. M. S. Rodarte (2014). *Grade of Membership: conceitos básicos e aplicação empírica usando o programa GoM para Windows, Linux e Stata*. População e Economia. Belo Horizonte: Editora UFMG.
- Guimarães, R. R. M., G. R. Guedes, and E. L. G. Rios-Neto (2010). De onde vim e até onde vou: uma análise preliminar da desigualdade socioeconômica e entrada no ensino superior brasileiro. *working paper*. belo horizonte: Cedeplar-ufmg.
- Hausman, J. A. and D. A. Wise (1977). Social experimentation, truncated distributions, and efficient estimation. *Econometrica: Journal of the Econometric Society*, 919–938.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153–161.
- Heckman, J. J., H. Ichimura, and P. E. Todd (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies* 64(4), 605–654.
- Heckman, J. J., J. Smith, and N. Clements (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies* 64(4), 487–535.
- Heckman, J. J. and J. A. Smith (1995). Assessing the case for social experiments. *The Journal of Economic Perspectives* 9(2), 85–110.
- Hull, J. R. and G. R. Guedes (2013). Reconstructing babel: comparando o bem-estar multidimensional de diferentes contextos culturais em busca de soluções comuns para o desenvolvimento. *Revista Brasileira de Estudos de População* 30(1), 271–297.
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*, Volume 7. Sage.
- Macmillan, R. and R. Copher (2005). Families in the life course: Interdependency of roles, role configurations, and pathways. *Journal of Marriage and Family* 67(4), 858–879.

- Macmillan, R. and S. R. Eliason (2003). Characterizing the life course as role configurations and pathways. *Handbook of the life course*, 529–554.
- Magidson, J. and J. K. Vermunt (2001). Latent class factor and cluster models, bi-plots, and related graphical displays. *Sociological methodology* 31(1), 223–264.
- Magidson, J. and J. K. Vermunt (2002). Latent class models for clustering: A comparison with k-means. *Canadian Journal of Marketing Research* 20(1), 36–43.
- Maddala, G. S. (1983). *Limited Dependent and Qualitative Variables in Econometrics*. Capítulo 9. Cambridge, England: Cambridge University Press.
- Manton, K. G., M. A. Woodbury, and H. D. Tolley (1994). *Statistical Applications Using Fuzzy Sets*. Wiley series in probability and mathematical statistics: Probability and mathematical statistics. John Wiley & Sons, Incorporated.
- Melo, F. L. B. (2007). Casais na grande são paulo: investigando a diversidade. *Nova economia* 17(2), 207–240.
- Miranda-Ribeiro, P., A. B. Simão, A. J. Caetano, I. H. O. Perpétuo, M. A. Lacerda, and M. E. A. Torres (2007). Acesso à contracepção e ao diagnóstico do câncer de colo uterino em belo horizonte: uma contribuição metodológica aos estudos quanti-quali. *Revista Brasileira de Estudos de População* 24(2), 341–344.
- Moisio, P. (2004). A latent class application to the multidimensional measurement of poverty. *Quality and Quantity* 38(6), 703–717.
- Reis, D. S. (2009). As redes para além dos rios: urbanização e desequilíbrios na amazônia brasileira. Tese (doutorado em demografia), Universidade Federal de Minas Gerais.
- Reis, D. S. and G. R. Guedes (2012). Análise urbana global: aplicação do modelo de classe latente para as *leading world cities*. *Geografia (Rio Claro. Impresso)* 37, 391–408.
- Sawyer, D. O., I. C. Leite, and R. Alexandrino (2002). Perfis de utilização de serviços de saúde no brasil. *Ciênc Saúde Coletiva* 7(4), 757–76.
- Vermunt, J. K. and J. Magidson (2001). Latent Class Cluster Analysis. In Hagenaars, J. A. and McCutcheon, A. L. (Eds.), *Applied Latent Class Analysis*. Cambridge, England: Cambridge University Press.
- Vermunt, J. K. and J. Magidson (2005a). *Technical Guide for Latent GOLD 4.0: Basic and Advanced*. Belmont Massachusetts: Statistical Innovations Inc.
- Vermunt, J. K. and J. Magidson (2005b). *User Guide for Latent GOLD 4.0*. Belmont Massachusetts: Statistical Innovations Inc.
- Woodbury, M. A. and J. Clive (1974). Clinical pure types as a fuzzy partition. *Journal of Cybernetics* 4(3), 111–121.
- Woodbury, M. A., J. Clive, and A. Garson Jr (1978). Mathematical typology: a grade of membership technique for obtaining disease definition. *Computers and biomedical research* 11(3), 277–298.
- Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Capítulo 17. Cambridge, MA: MIT Press.