

Taller CEPAL – 21-22 Novembro 2018

Estimação de Indicadores Sociais para ‘Áreas Pequenas’ no Brasil

Pedro Silva

Roteiro

Alguns exemplos de aplicação de métodos de **Estimação** **Para Áreas Pequenas (EPAP)** no Brasil

Censos Demográficos

Fecundidade

Mapa de pobreza 2008

Mortalidade infantil

Acesso à internet e posse de
computador nos domicílios

Totais por atividades nos
serviços

População total, e por sexo × idade

Dificuldades e Barreiras

Oportunidades e Perspectivas

Aplicações Pioneiras

Censos Demográficos decenais no Brasil usam amostragem para coleta de dados socioeconômicos desde 1960.

Amostras são **estratificadas por setor censitário**, e **conglomeradas em 1 estágio** (sorteio de 1 a cada K domicílios por setor). K varia entre: 2; 3; 4; 5; 10; 20.

Todos os moradores pesquisados nos domicílios selecionados.

Dados dessas amostras eram usados para estimar para áreas pequenas, tais como **pequenos municípios**, e **áreas de ponderação** em municípios maiores.

Aplicações Pioneiras

Método de estimação principal era **calibração dos pesos** para totais populacionais apurados nos censos.

No Censo de 1980, **Tabulações Avançadas** foram produzidas usando **estimadores sintéticos** – [IBGE \(1982\)](#); [Garcia \(1986\)](#).

Comparando com atual sofisticação metodológica, tais **aplicações** podem ser classificadas como **muito simples**.

Porém abriam caminho para a obtenção de EPAPs ou usando pequenas amostras **antes** do processamento completo dos dados dos censos.

Censos Demográficos

Tradição continuou nos censos recentes: dados da amostra são tabulados por **áreas de ponderação**, definidas para ter tamanho mínimo igual a 400 domicílios na amostra.

Não têm sido fornecidas estimativas para áreas menores que as **áreas de ponderação** no Censo brasileiro.

No Censo Demográfico de 2010, Brasil foi subdividido em **10.184 áreas de ponderação**, sendo 4.443 delas iguais a municípios inteiros - [IBGE \(2016\)](#).

Mapa de Pobreza 2008

IBGE (2008) publicou um **Mapa de Pobreza por Municípios**, usando método para EPAP proposto por [Elbers, Lanjouw e Lanjouw \(2003\)](#).

Ideia central: ajustar modelo para explicar Despesa Mensal Domiciliar Per Capita (DMDPC) usando dados da **Pesquisa de Orçamentos Familiares (POF) 2002/03**.

Depois usar modelo ajustado na POF para prever DMDPC com dados domiciliares individuais da **Amostra do Censo 2000** e então **estimar indicadores de pobreza por municípios**.

Mapa de Pobreza 2008

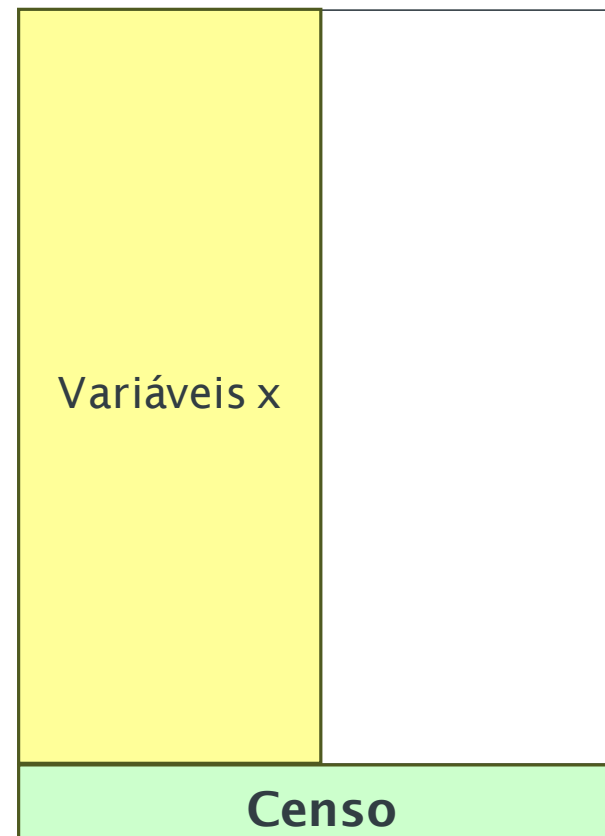
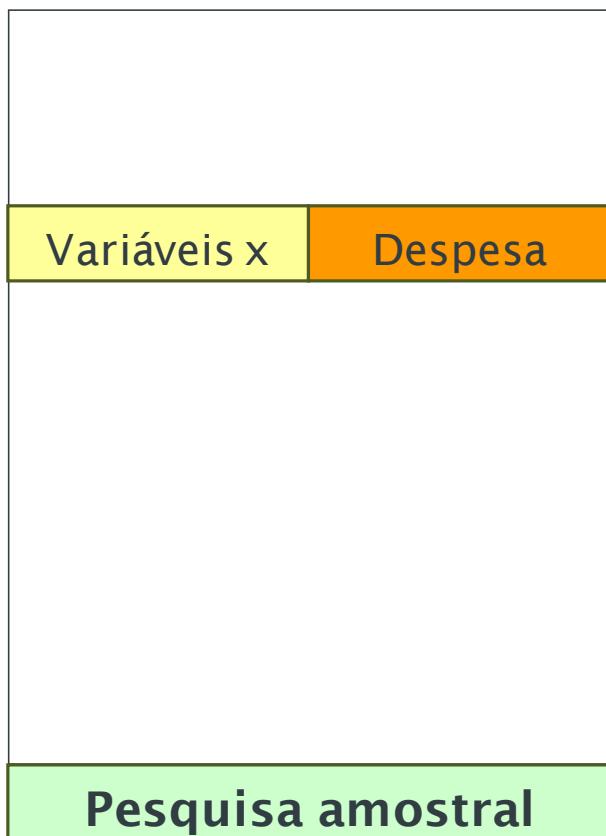
Pesquisa orçamentos familiares

- Despesas medidas com alta qualidade
- Boa informação auxiliar
- Amostra pequena, desenhada para níveis agregados (50k)
- Incapaz de estimar com precisão para áreas pequenas
- Mais atual

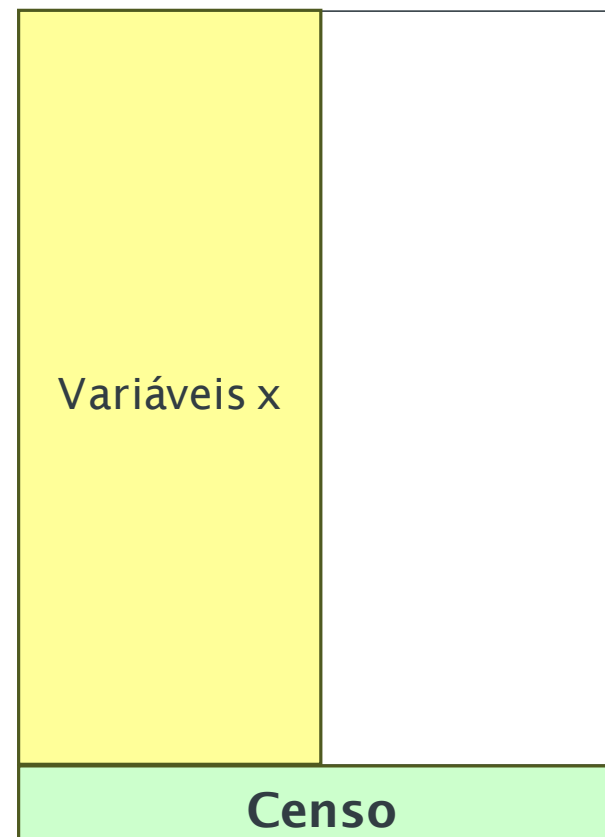
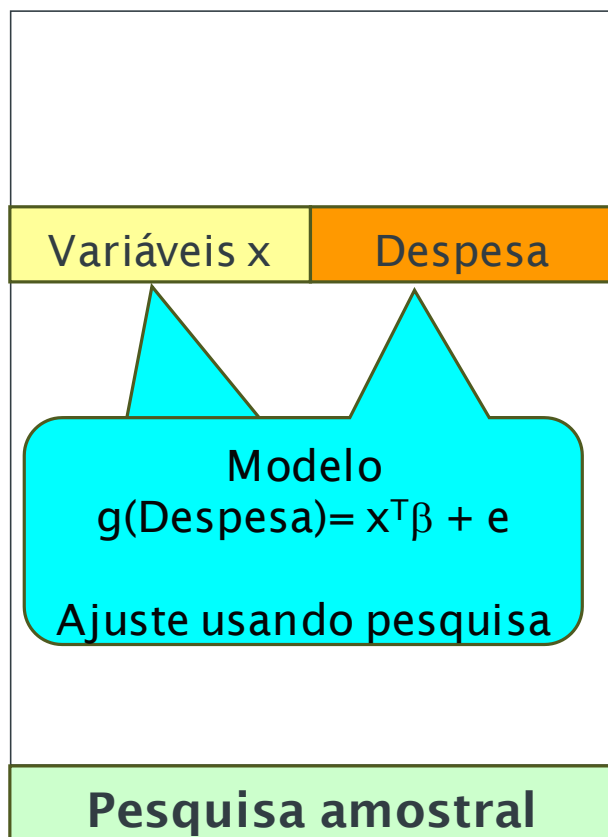
Censo

- Não têm dados de despesas
- Boa informação auxiliar
- Amostra muito grande (6.2m)
- Estimação para geografia detalhada
- Menos atual

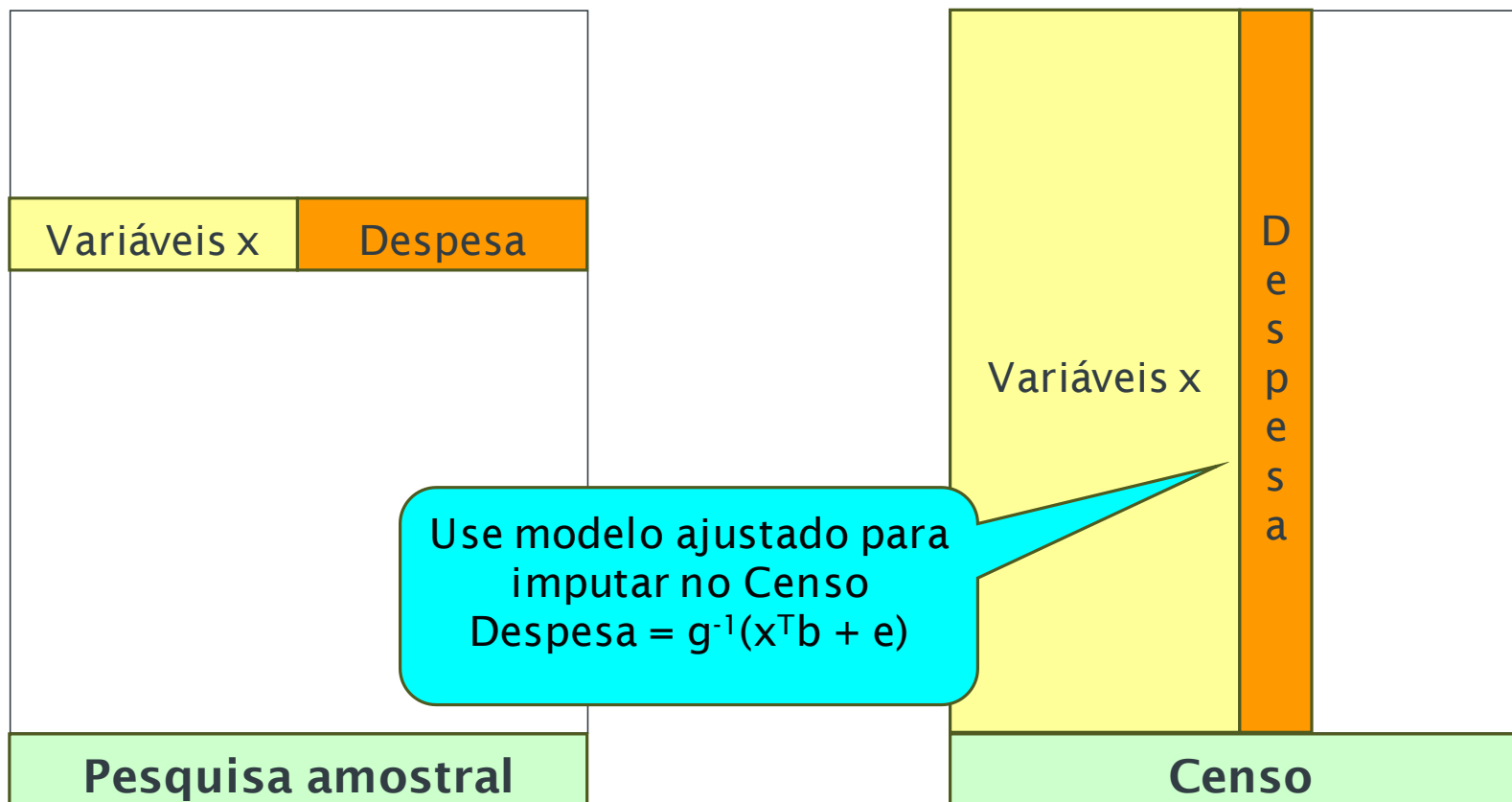
Mapa de Pobreza 2008



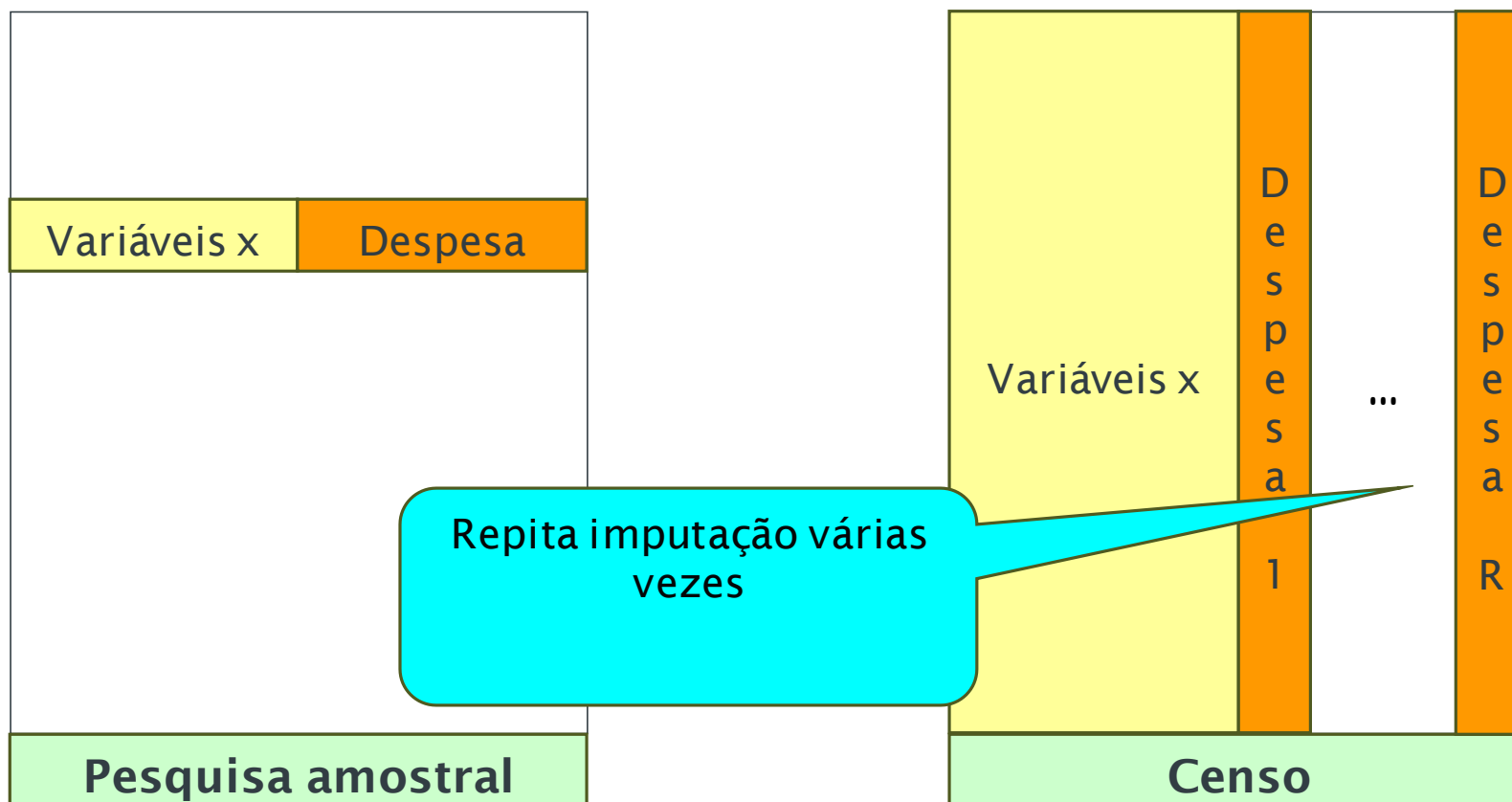
Mapa de Pobreza 2008



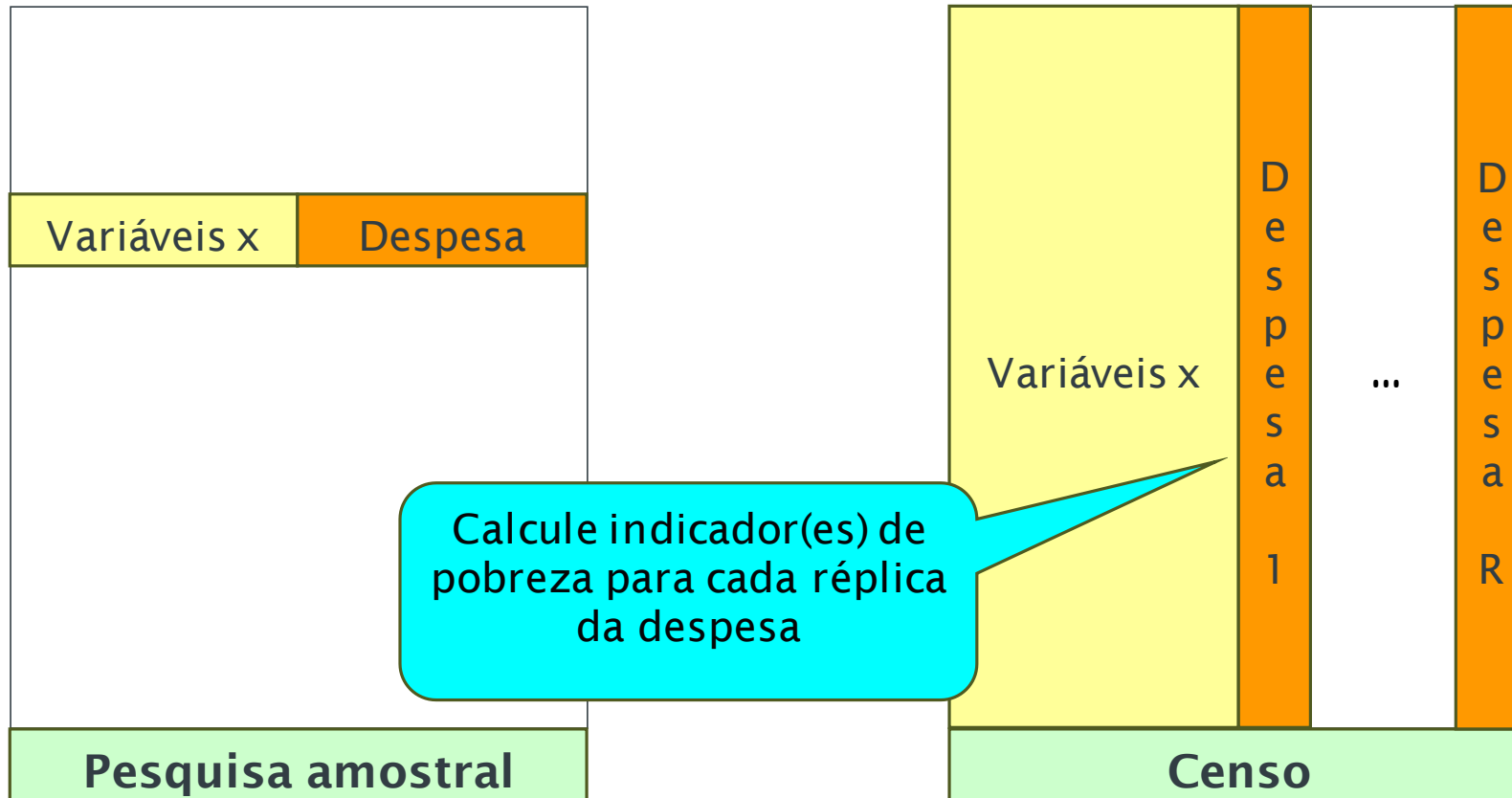
Mapa de Pobreza 2008



Mapa de Pobreza 2008



Mapa de Pobreza 2008



Mapa de Pobreza 2008

Avaliação das estimativas conduzidas usando Censo Demográfico 2000 – [Cunha \(2007\)](#); [Antonaci, Silva e Moura \(2013\)](#).

Dados de **RMDPC** usados para definir indicadores de pobreza.

Estado de Minas Gerais usado na avaliação: 851 municípios, dos quais apenas 150 tinham amostra na POF 2002/03.

Estimativas EPAP da **% de domicílios pobres** comparadas com estimativas diretas obtidas da amostra do Censo.

Figura 1 – % de domicílios pobres por município

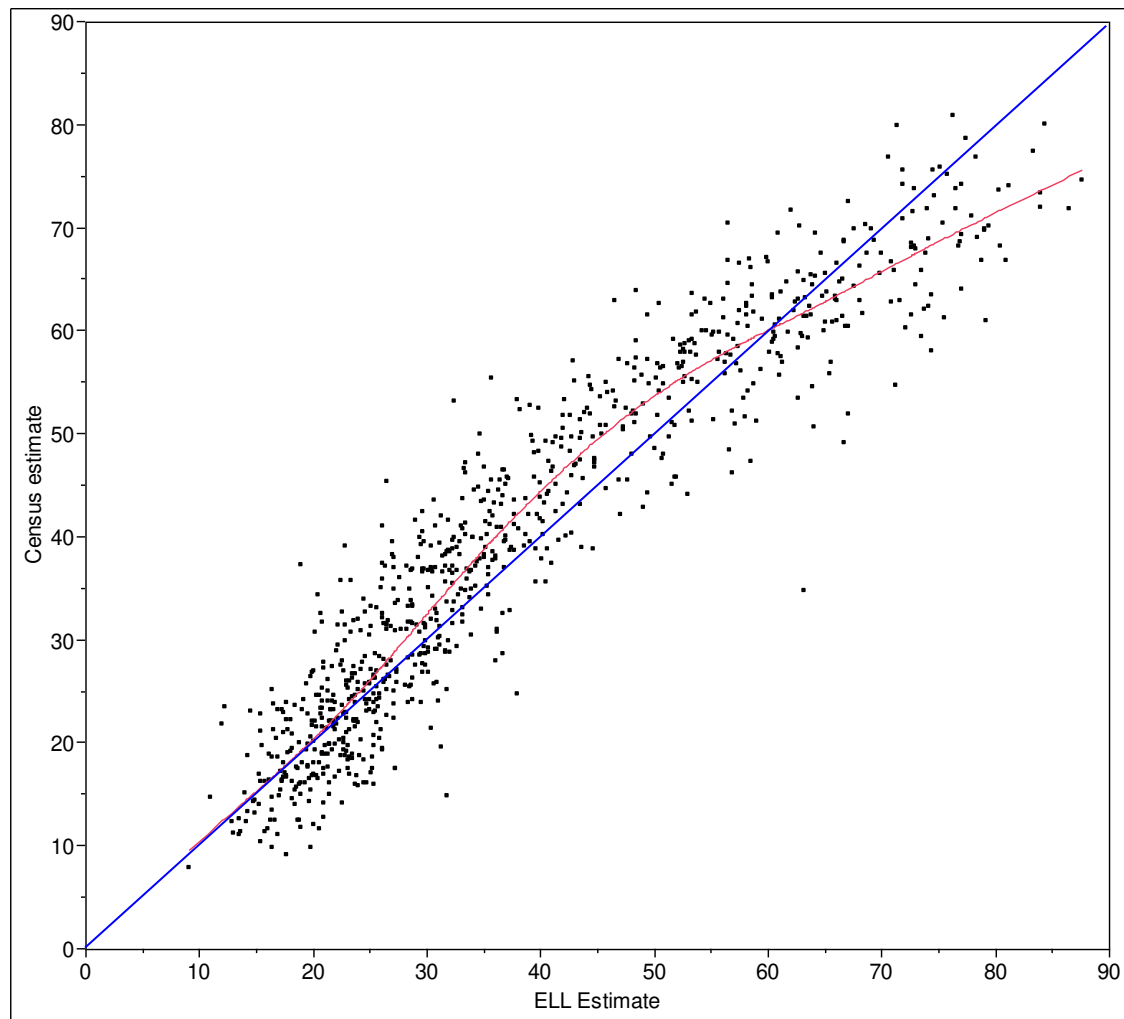
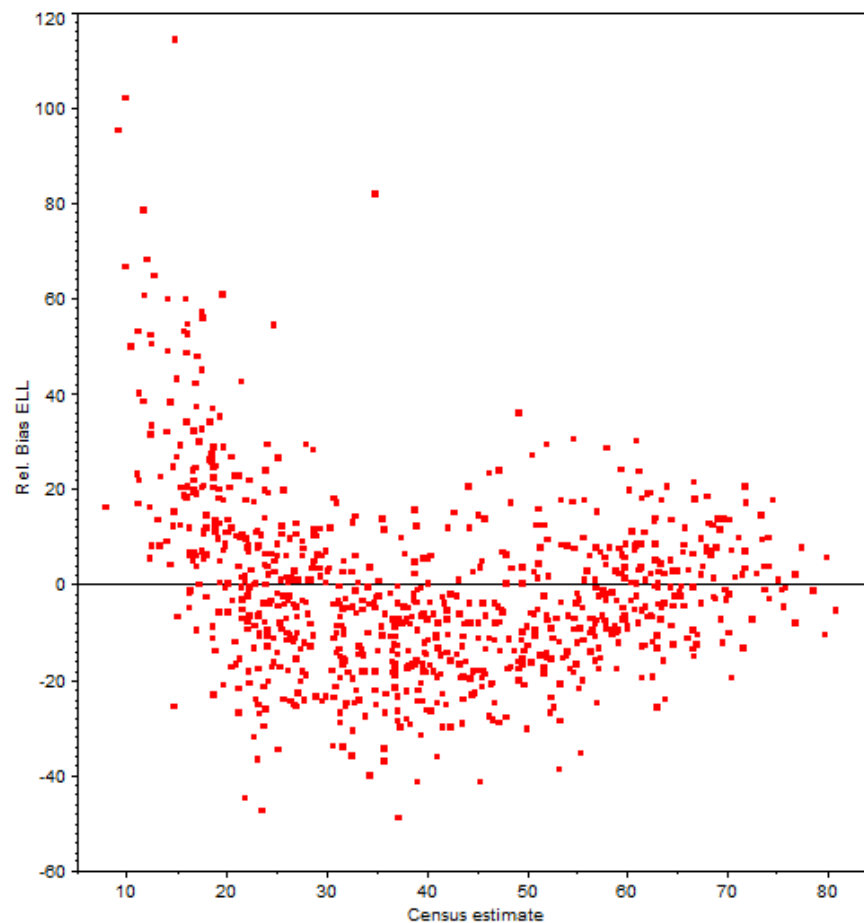
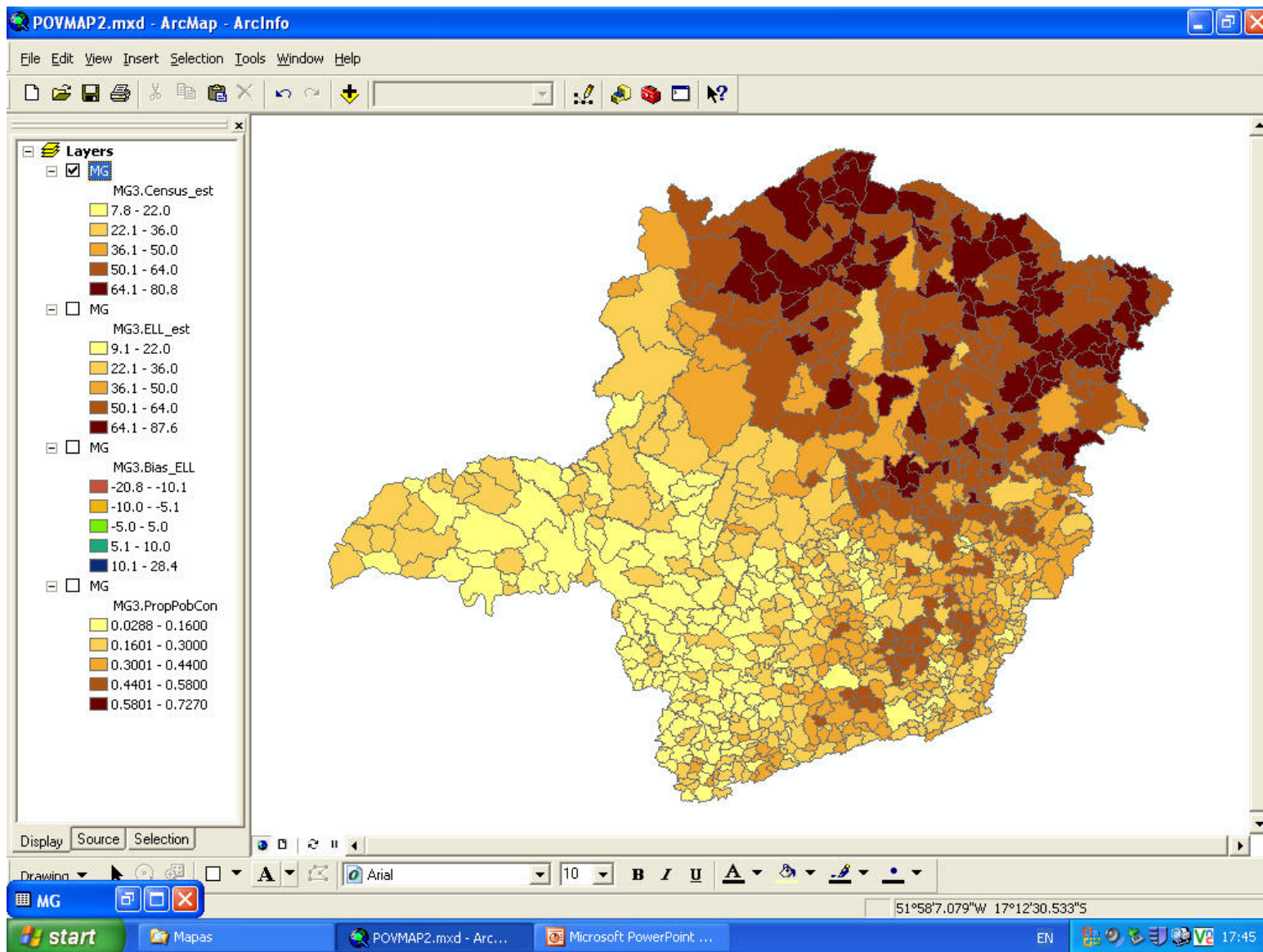
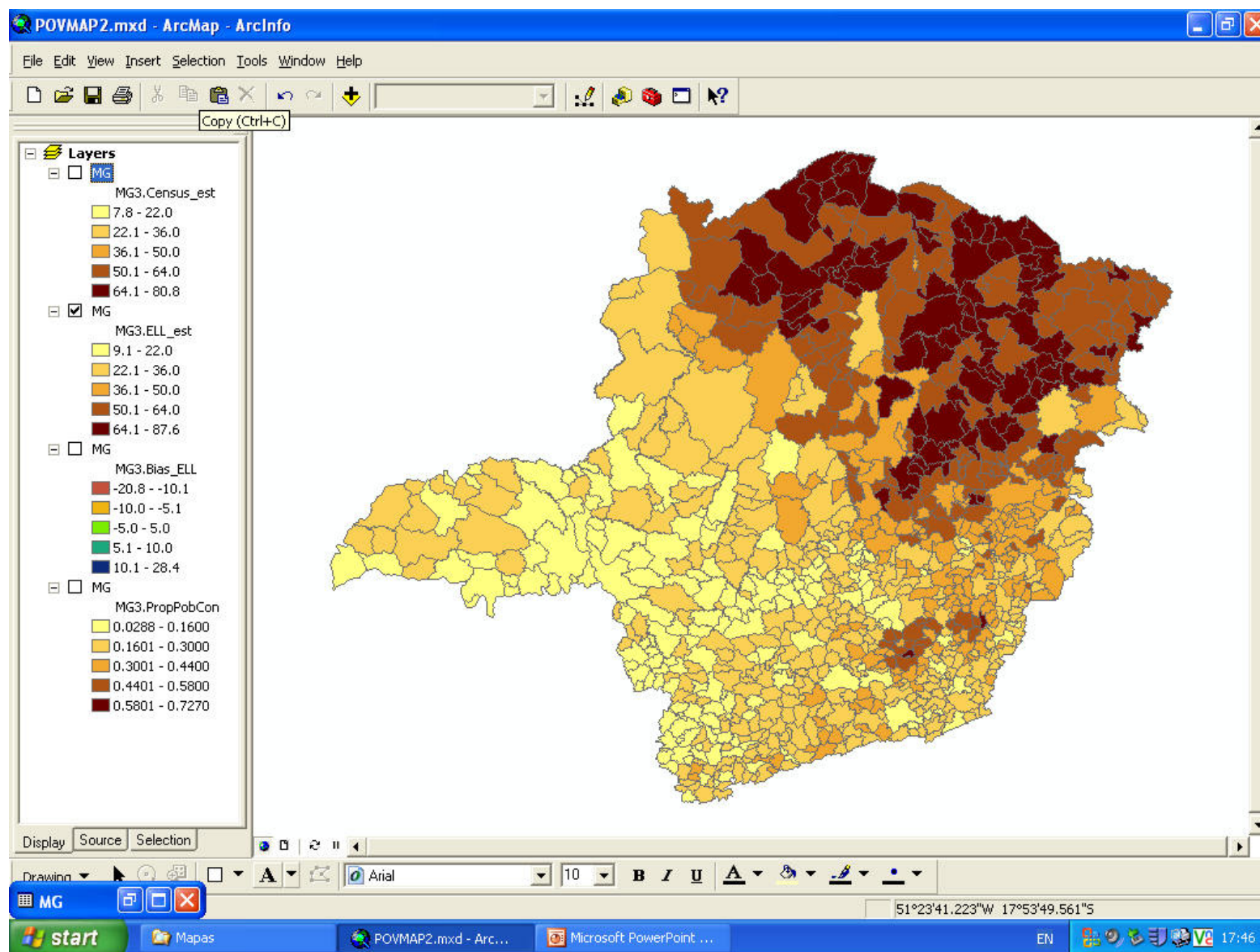


Figura 2 – Vício relativo das estimativas







Mapa de Pobreza 2008

Dificuldades:

- Dados da POF são **pouco frequentes** (2002/03; 2008/2009; 2017/18).
- Estimativas parecem ter **viés**;
- Métodos usados foram logo superados – [Molina & Rao \(2009, 2010\)](#);
- Computação pesada (à época).

Resultado: experiência ainda não foi repetida nesta década.

Acesso à Internet e Posse de Computador

Pesquisa **TIC-Domicílios** realizada anualmente pelo NIC.br.

Amostra nacional de $\cong 20.500$ domicílios, planejada para dar resultados por macrorregiões do Brasil (NO, NE, SE, SU, CO).

Interesse: produzir e publicar estimativas por Unidades da Federação (27).

Problema: amostra é pequena em estados das regiões NO e CO, por razões de custo.

Cavalcante (2016); Coelho et al (2018).

Acesso à Internet e Posse de Computador

Estudo com métodos de estimação de áreas pequenas:

- Usou dados das pesquisas de 2014 e 2015;
- Três métodos:
 - Estimador composto intra-anual;
 - Acumulação de amostras de 2 anos (*'pooling'*);
 - Acumulação de amostras de 2 anos + estimador composto.
- Avaliação das estratégias por comparação com estimativas da PNAD 2015, e de estimativas do RMSE.

Acesso à Internet e Posse de Computador

Estimador composto intra-anual

$$\hat{\theta}_d^C = \varphi_d \hat{\theta}_d + (1 - \varphi_d) \hat{\theta}_{R(d)}$$

$\hat{\theta}_d$ é estimativa direta para UF d ;

$\hat{\theta}_{R(d)}$ é estimativa direta para macrorregião da UF d ;

φ_d é o peso da estimativa direta da UF d .

Novidade: pesos derivados considerando indicadores R da representatividade da amostra de cada UF.

Acesso à Internet e Posse de Computador

Estimador usando acumulação de amostras

$$\hat{\theta}_d^A = \frac{1}{2} [\hat{\theta}_d^{t-1} + \hat{\theta}_d^t]$$

$\hat{\theta}_d^t$ é estimativa direta para UF d no ano t .

Estimador composto usando acumulação de amostras

$$\hat{\theta}_d^{AC} = \varphi_d \hat{\theta}_d^A + (1 - \varphi_d) \hat{\theta}_{R(d)}^A$$

$\hat{\theta}_{R(d)}^A$ é estimativa para macrorregião da UF d obtida combinando amostras de dois anos.

Acesso à Internet e Posse de Computador

Resultado: como esperado, método combinando amostras de dois anos com estimador composto parece ter fornecido melhores estimativas.

Estimativas ainda não publicadas como resultado ‘oficial’.

Plano é aplicar abordagem a partir da pesquisa de 2017, com dados retroativos para alguns anos.

Desafios ainda pendentes:

- Estimação do EQM;
- Validação externa.

Estimativas Populacionais por Municípios

Assunção (2002) comparou diversos métodos para projetar / estimar **população total por municípios** no Brasil.

Método Bayesiano Espacial proposto no trabalho mostrou resultados superiores aos métodos em uso pelo IBGE e outros com os quais foi comparado.

Infelizmente, **método não foi adotado** para produção das estimativas municipais de população pelo IBGE.

Motivos principais: metodologia sofisticada; software genérico não disponível; equipes responsáveis inseguras.

Estimativas Populacionais por ‘Bairros’

Pitta, Dini e Jannuzzi (2010) produziram estimativas populacionais anuais, para o período de 2001 a 2010, por sexo e grupos etários, para cerca de 400 áreas pequenas (‘bairros’) definidas para o município de São Paulo.

Modelaram proporções da população total em cada ‘bairro’, usando informações auxiliares de ‘variáveis sintomáticas’.

Porém estimativas populacionais nunca foram publicadas como estatísticas oficiais.

Fecundidade por Municípios

Cavenaghi et al. (2004) + Assunção et al. (2005) obtiveram estimativas das **taxas brutas de fecundidade por municípios** para o Brasil.

Usaram **método Bayesiano Empírico** para produzir estimador composto, onde estimativa direta do município era combinada com estimativa de uma vizinhança local do município.

Método foi avaliado como bem sucedido, porém nunca foi adotado na elaboração de estimativas oficiais da fecundidade nos municípios.

Mortalidade Infantil por Municípios

Silva (2015) estimou **taxas de mortalidade infantil** por municípios do estado de Pernambuco usando **Modelo Linear Generalizado Misto**.

Dados de registros administrativos, mas sujeitos a variação estocástica por causa eventos raros das pequenas populações.

Variáveis preditoras consideradas provenientes dos Censos Demográficos e fontes administrativas.

Modelo ajustou-se bem aos dados.

Método parece superar a estratégia habitual de usar médias móveis de três anos para suavizar as taxas municipais.

Dados sobre Serviços por Atividade

Moura, Neves e Silva (2017) estimaram totais de variáveis econômicas usando dados da Pesquisa Anual de Serviços para 104 domínios definidos por UF x Atividade.

Método usado é no nível de domínio, e se baseia na distribuição normal assimétrica, pois dados têm forte assimetria.

Resultados promissores: método proposto reduziu os coeficientes de variação em cerca de 92,3% dos domínios.

Pesquisa continua: um dos autores é estudante de doutorado na ENCE...

Estado da Arte no Brasil

Não há **aplicações regulares** de métodos para EPAP na produção de estatísticas oficiais.

Pressões têm aumentado para:

- Produção de estimativas mais detalhadas para monitorar indicadores dos ODS;
- Contenção de custos ou redução de recursos à disposição das agências produtoras de informações oficiais;
- Ampliação do uso de novas fontes de dados para produção de informações.

Dificuldades e Barreiras no Brasil

Métodos para EPAP dependem de:

- Modelos sofisticados;
- Dados auxiliares de boa qualidade, e correlacionados com variáveis de interesse.

Escassez de software e métodos ‘genéricos’.

Capacidade estatística limitada nas agências produtoras.

Falta de **estudos de avaliação conclusivos** de esforços já feitos para obter e publicar estimativas para áreas pequenas.

Produtores de estatísticas oficiais **‘avessos a risco’**.

Oportunidades no Brasil

Aparente contradição representa existência de **oportunidades** para aplicações de EPAP no Brasil.

Capacitação técnica de equipes tem aumentado.

Ferramentas de software ‘genéricas’ para EPAP:

[sae \(Molina e Marhuenda, 2015\)](#)

<http://www.sample-project.eu/images/stories/deliverables>

<http://www.bias-project.org.uk/software/#sae>

Protocolo Rotineiro de Avaliação Externa (PRAE)

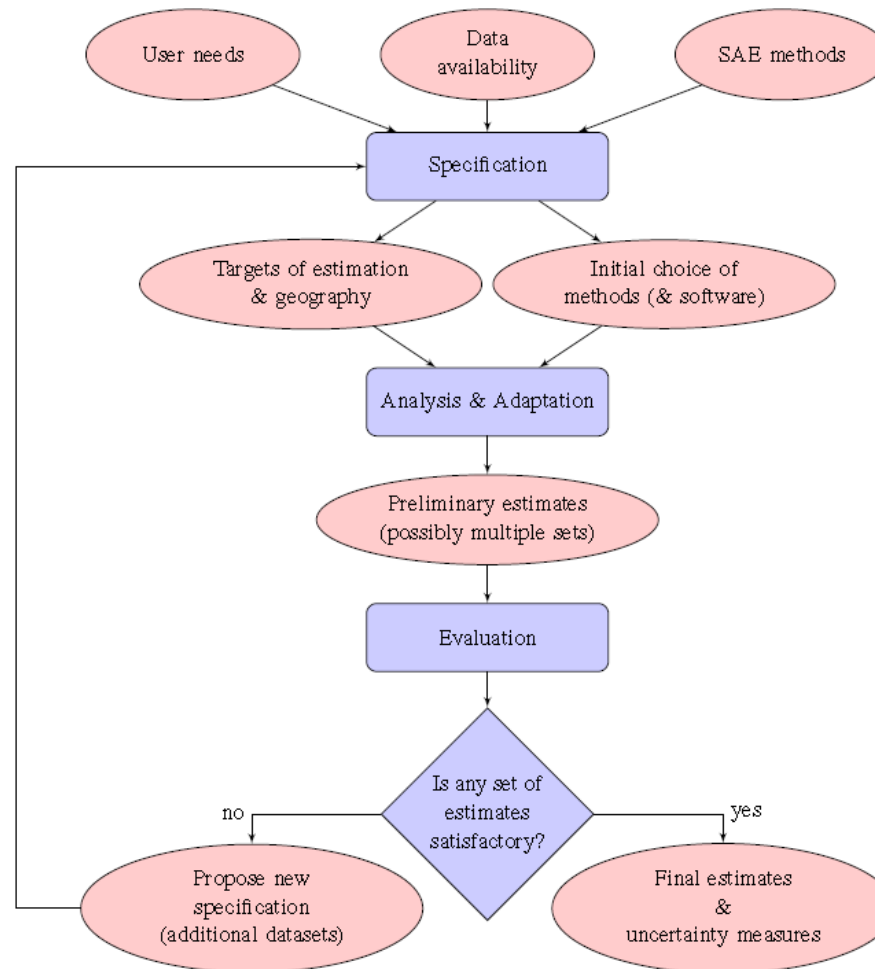
Dorfman (2018) propõe adoção do PRAE como forma de proteção contra aplicações insatisfatórias da EPA.

“Current practice in small area estimation makes us vulnerable to our using very elegant and persuasive techniques that leave us in the dark as to whether they are actually working in the particular survey to which they are applied. This is a serious matter, especially because small area estimates are often used to make judgments on funding and other matters important to the body politic.”

Protocolo Rotineiro de Avaliação Externa (PRAE)

1. Para cada situação de pesquisa em que EPAP vá ser necessária, uma **amostra independente** de áreas deve ser extraída para ser usada na avaliação das EPAPs.
2. Nessas áreas, amostras ‘locais’ de tamanho suficiente para produzir **estimativas diretas** de boa qualidade devem ser pesquisadas.
3. Dados dessas amostras devem ser então usados para avaliar a qualidade das EPAPs usando medidas de diagnóstico propostas.
4. Ao final, amostras locais podem ser combinadas à amostra inicial para aprimorar estimativas.

Protocolo Para Aplicações (Tzavidis et al. 2018)



Obrigado por sua atenção.

Referências

Antonaci, Giuseppe; Silva, P. L. d. N.; Moura, F. A. S.. A comparison of methods to estimate poverty indexes in small samples. Hong Kong: Proceedings of the World Statistics Congress, International Statistical Institute, 2013.

Assunção, R.M.; Schmertmann, C.P.; Potter, J.E.; Cavenaghi, S.M. Empirical Bayes estimation of demographic schedules for small areas. *Demography*. v. 42, n. 3, p. 537-558. 2005.

Barbosa e Silva, C. J.. Comportamento da mortalidade infantil e suas componentes etárias nos municípios de Pernambuco. Escola Nacional de Ciências Estatísticas, 2015. 173 p.

Cavalcante, C. V.. Estimação em Pequenas Áreas para o Acesso às Tecnologias de Informação e Comunicação na Pesquisa TIC Domicílios. Escola Nacional de Ciências Estatísticas, 2016. 119 p.

Cavenaghi, S. et al. Estimating Total Fertility Rates for Small Areas in Brazil. XIV Encontro Nacional de Estudos Populacionais, ABEP, Caxambu, 2004. 28 p.

Coelho, I. B.; Pitta, M. T.; Silva, P. L. d. N.. Estimates for Information and Communication Technologies State Level Indicators from ICT Household Surveys in Brazil. São Paulo, Brazil, 2018.

Cunha, M. B. A. M. D.. Mapas de Pobreza: Avaliação da Metodologia e de sua aplicação para retratar a pobreza no Rio Grande do Norte. Escola Nacional de Ciências Estatísticas, 2007. 160 p.

Dorfman, A. H. Towards a Routine External Evaluation Protocol for Small Area Estimation. International Statistical Review v. 86, n. 2, p. 259–274, 2018.

Elbers, C.; Lanjouw, J. O.; Lanjouw, P.. Micro–Level Estimation of Poverty and Inequality. Econometrica v. 71, n. 1, p. 355–364, 2003.

Garcia, R. M.. *Estimação em pequeno domínio*. Rio de Janeiro, 1986. 66 p. Instituto de Matemática Pura e Aplicada – IMPA.

Ghosh, M.; Myung, J.; Moura, F. A. S. *Robust Bayesian small area estimation*. *Survey Methodology* v. 44, n. 1, p. 101–115, 2018.

IBGE. *Tabulações avançadas do Censo Demográfico*. Rio de Janeiro, Brasil: IBGE, 1982. 160 p.

IBGE. *Mapa de pobreza e desigualdade: municípios brasileiros 2003*. Rio de Janeiro, Brasil: IBGE, 2008.

IBGE. *Metodologia do Censo Demográfico 2010, 2a edição*. Rio de Janeiro, Brasil: Instituto Brasileiro de Geografia e Estatística, 2016. 711 p.

Molina, I.; Marhuenda, Y. *sae: An R Package for Small Area Estimation*. *The R Journal* v. 7, n. 1, p. 81–98, 2015.

Molina, I.; Rao, J. N. K. *Small Area Estimation of Poverty Indicators*. Small Area Estimation Conference. Elche, Spain, 2009.

Molina, I.; Rao, J. N. K. *Small area estimation of poverty indicators*. *The Canadian Journal of Statistics*, v. 38, n. 3, p. 369–385, 2010.

Moura, F. A. S.; Neves, A. F.E.; Silva, D. B. D. N. . Small area models for skewed Brazilian business survey data. *Journal of the Royal Statistical Society - Series A - Statistics In Society*, v. 180, p. 1039-1055, 2017.

Pitta, M. T.; Dini, N. P.; Jannuzzi, P. D. M.. Modelos para produção de estimativas demográficas para bairros e domínios intraurbanos: a experiência para o município de São Paulo. XVII Encontro Nacional de Estudos Populacionais, ABEP, Caxambu - MG 13 p., 2010.

Tzavidis, N. et al. From start to finish: a framework for the production of small area official statistics. *Journal of the Royal Statistical Society. Series A: Statistics in Society* v. 181, n. 4, p. 927–979, 2018.