

Pesquisa longitudinal a partir de amostra selecionada do censo 2020

Pedro Silva

ENCE – Escola Nacional de Ciências Estatísticas

Sistema de pesquisas populacionais do IBGE

Baseado em duas **coleções de pesquisas domiciliares + registro administrativo:**

Censo demográfico decenal

- Censo tradicional +
- Pesquisa socioeconômica por amostragem

Pesquisas por amostragem do SIPD

- Repetidas: PNAD Contínua
- Transversais: POF; PNS; etc.

Estatísticas do Registro Civil

Estudos (Amostrais) Transversais

Aplicáveis para:

- Medir o **nível** de indicadores em níveis agregados;
- Descobrir e revelar **associações** entre variáveis;
- Geralmente **observacionais (não experimentais)**.

Não aplicáveis para:

- Identificar relações de **causa e efeito**;
- Fornecer resultados para **domínios muito detalhados**.

Mais baratos.

Mais frequentes.

Estudos Longitudinais

Requeridos para:

- Medir e avaliar **variações em nível individual**;
- Tentar identificar relações de **causa e efeito**;
- **Avaliações de impacto** de certas intervenções.

Mais caros.

Mais raros:

- **Custos** substanciais envolvidos;
- Planejamento e compromissos de **prazos + longos**;
- **Complexos** de planejar e analisar.

Censo decenal

‘**Retrato completo**’ da população, porém:

- ‘**Alta resolução**’ para **poucas variáveis**: contagens da população para **muitos pequenos domínios** (setores; grade; bairros;).
- ‘**Média resolução**’ para conjunto de variáveis do **questionário longo**: estimativas de indicadores socioeconômicos para **domínios geográficos médios** (áreas de ponderação; municípios).
- **Baixa frequência.**

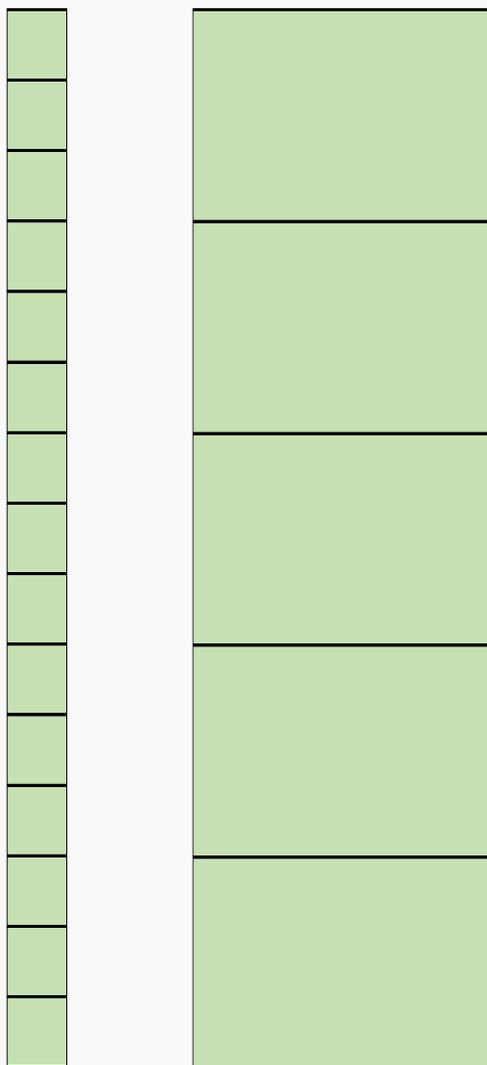
Censo Decenal



Censo + Amostra Decenal

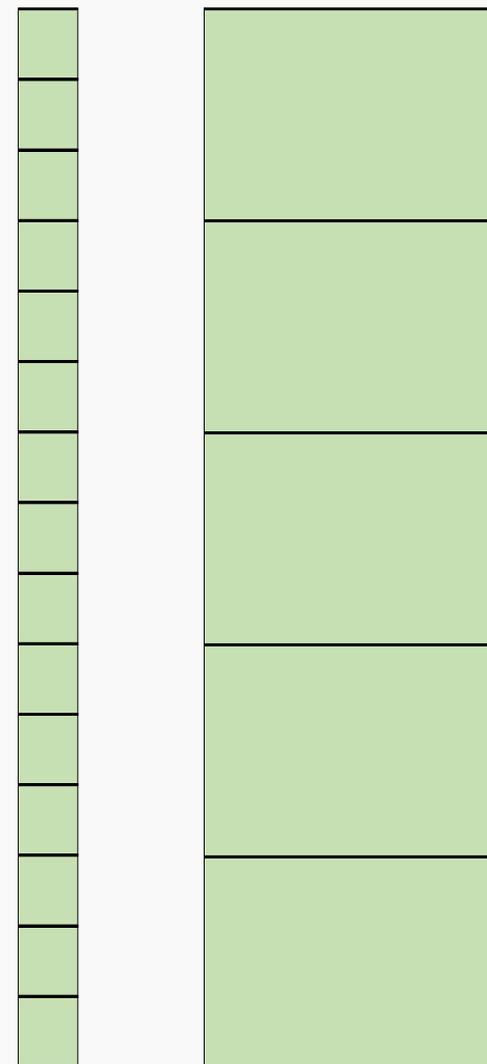
Ano t

Censo Amostra



Ano t+10

Censo Amostra



Pesquisas Domiciliares do SIPD

‘**Retratos + frequentes**’ da população, porém:

- ‘**Baixa resolução**’ para **muitas variáveis**: estimativas de amplo conjunto de indicadores socioeconômicos da população para **poucos grandes domínios** (UFs; RMs);
- Oportunidades para monitorar **variações brutas** de indicadores ao longo do tempo (frequência mensal, trimestral ou anual);
- Oportunidades para investigar **temas adicionais** a cada ocasião;
- Seguimento de domicílios / pessoas apenas por **curtos períodos de tempo** na PNADC (e na extinta PME).

Pesquisas Amostrais Repetidas

t	t+1	t+2	t+3

Estatísticas do Registro Civil

‘**Registros de eventos vitais**’ da população via cartórios:

- ‘**Alta resolução**’ para **poucas variáveis**;
- ‘**Alta frequência**’, mas dados obtidos retrospectivamente;
- **Coleta trimestral**, com **divulgação anual** de estatísticas resumo;
- **Cobertura desigual** no território nacional;
- **Registro tardio** de eventos;
- Dados obtidos de muitas fontes, com **qualidade variável**.

Outros Registros de Eventos Vitais

SINASC - Sistema de Informações de Nascidos Vivos.

SIM - Sistema de Informações de Mortalidade.

‘Registros de eventos vitais’ da população **via estabelecimentos de Saúde:**

- **‘Alta resolução’** para **poucas variáveis;**
- **‘Alta frequência’** (dados obtidos continuamente);
- **Informações adicionais** sobre eventos;
- **Coleta contínua** com divulgação anual;
- **Cobertura desigual** no território nacional;
- **Formulários** de registro **padronizados.**

Fraquezas do Sistema de pesquisas populacionais do IBGE

Nenhuma integração entre as fontes domiciliares e os registros vitais, nem mesmo os processados pelo IBGE.

Nenhuma integração dos registros vitais processados pelo IBGE com os da área da Saúde.

Custos de reputação por discrepâncias entre informações de diferentes fontes.

Custos de duplicação e sobrecarga de coleta por redundância de coleta de informações sobre eventos vitais.

Oportunidades perdidas quanto às possibilidades de combinação de informações das distintas fontes.

Como podemos melhorar o sistema?

O Contexto Atual

Ferramentas computacionais e métodos disponíveis favorecem possibilidades de estudos baseados em **pareamento e combinação de fontes de dados**.

Fontes sobre **eventos vitais** dispõem de dados com **boa cobertura e qualidade**.

Dados de fontes orgânicas também disponíveis para complementar fontes tradicionais.

Oportunidade de planejar estudo com amostragem controlada.

Oportunidade para **complementar portfolio** de sistema de pesquisas populacionais do IBGE.

Sistema de pesquisas populacionais do IBGE

Baseado em **três** coleções de pesquisas domiciliares + registros vitais:

- ❑ Censo demográfico decenal
- ❑ Pesquisas por amostragem do SIPD
- ❑ Estatísticas de registros vitais
- ❑ **Pesquisa Longitudinal do IBGE**
 - Similar à do Reino Unido;
 - Planejada e iniciada a partir do CD 2020;
 - Combinando dados de censos e dos registros vitais.

Pesquisa Longitudinal do Reino Unido (LS)

Estudo contém dados pareados do **censo decenal** e **eventos vitais** para uma **amostra de 1% da população** da Inglaterra e País de Gales.

Iniciada em 1974, com amostra da **população residente** nos países cobertos, conforme enumerados no Censo de 1971 (ref. 25/04/1971).

No momento, dados de **5 censos** (1971 – 2011) + **eventos vitais** ao longo de **40 anos**.

Ver por exemplo: Hattersley, L., & Creeser, R. (1995).

LS - Motivação principal

Capacidade de seguir amostra selecionada por longo período de tempo (longitudinal ‘longo’);

Dados para aprimorar análises sobre **mortalidade relacionada com a situação ocupacional**; e

Dados mais detalhados sobre **fecundidade e espaçamento de nascimentos**;

Análises sobre migrações e outros temas.

LS - Amostragem

Amostra foi baseada em selecionar todos os indivíduos nascidos em **quatro dias** equidistantes no ano.

Por exemplo, suponha que o dia 14 do mês 3 foi escolhido para partida.

Então a amostra seria composta por todos os nascidos em **14/3, 13/6, 12/9 e 12/12 de cada ano.**

Domicílios onde residiam tais indivíduos e todos os seus moradores também selecionados a cada censo, para permitir análises transversais adequadas.

Mas apenas indivíduos selecionados são seguidos ao longo do tempo.

LS - Metodologia

Planejamento corresponde a um **estudo multi-coortes, prospectivo, contínuo**.

Amostras de atualização são selecionadas a cada censo usando o mesmo método, e depois agregadas à pesquisa.

Entre os censos indivíduos entram por **nascimento** ou **imigração** (se nascidos nos dias selecionados), e saem por **óbito** ou **emigração**.

Logo atualizações entre os censos são baseadas em registros vitais e administrativos.

→ **Pareamento** dos dados das várias fontes requerido.

LS - Vantagens

Amostra inicial tinha mais de **meio milhão** de pessoas;

Tamanho da amostra se mantém ou cresce ao longo do tempo, se população em crescimento;

Fontes de dados são censos e registros vitais, portanto de custo adicional de obtenção relativamente baixo;

Métodos de **pareamento** evitam problemas de não resposta comuns em estudos longitudinais prospectivos;

Possibilidades de uso para **estudos da cobertura** do censo e de registros vitais.

LS - Desvantagens / Dificuldades

Requer coleta da **data de nascimento** nos censos;

Custo de operações de pareamento;

Dificuldades com pareamento podem resultar em ‘**viés de seleção**’;

Amostra de poucos dias sujeita a risco de viés de amostragem;

Complexidade das análises;

Compromisso com manutenção da pesquisa por longo período de tempo.

LS – Tipos de Análises Possíveis

Análises de **dados pareados de censos com eventos vitais**.

Análises de **dados pareados de censos sucessivos**.

Análises de **dados pareados de eventos sucessivos**.

Análises **transversais** reunindo dados de registros administrativos pareados com os de um censo específico.

LS – Cuidados Requeridos

Proteção de confidencialidade dos microdados tem **prioridade máxima**.

Mecanismos de **armazenamento e controle de acesso** do mais alto nível de segurança.

Mecanismos para obtenção de acesso são **formais** e impõem **padrões éticos exigentes**.

Extrações limitadas aos conteúdos, usos e fins especificados nos requerimentos de acesso.

Controles rigorosos dos tipos de resultados que podem ser ‘exportados’ dos centros de acesso.

Pesquisa Longitudinal do IBGE × LS

Importante coletar dados com questionário longo (amostra) para amostra toda da PL no Censo;

Amostra total > 2 milhões, se mantida mesma fração (1%);

Maiores dificuldades para pareamento com registros vitais é esperada;

Investimento necessário em capacitação pode levar tempo;

Compromissos com **financiamento de longo prazo** mais difíceis de assegurar;

Oportunidade para planejar a primeira amostra ANTES de fazer o Censo 2020.

**Como podemos
viabilizar esta ideia?**

Obrigado por sua atenção.

www.ibge.gov.br

www.ence.ibge.gov.br

pedro-luis.silva@ibge.gov.br

Referências

Freitas, M. P. S. de, & Antonaci, G. (2014). *Sistema Integrado de Pesquisa Domiciliares Amostra Mestra 2010 e Amostra da PNAD Contínua*. Rio de Janeiro, Brasil.

Hattersley, L., & Creeser, R. (1995). *Longitudinal Study 1971-1991: History, organisation and quality of data. Series LS No. 7*. London: UK.

IBGE. (2014). *Pesquisa Nacional por Amostra de Domicílios Contínua - Notas Metodológicas - volume 1*. Rio de Janeiro, Brasil. IBGE(2016). *Estatísticas do Registro Civil 2015*. Rio de Janeiro: IBGE, Vol. 42.

Lynch, K., Leib, S., Warren, J., Rogers, N., & Buxton, J. (2011). *Longitudinal Study 2001 - 2011 Completeness of census linkage. Series LS No . 11*. London.