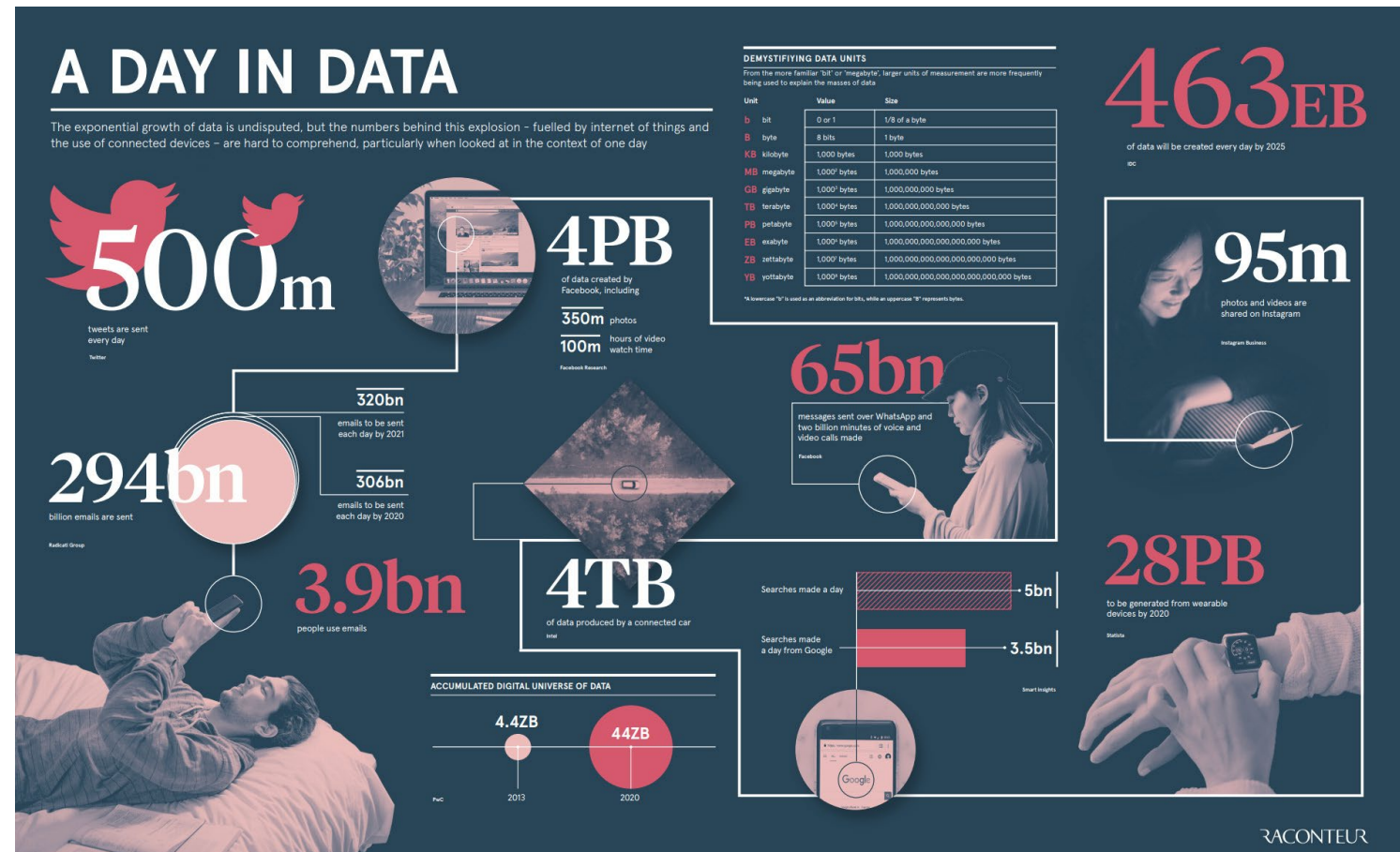


Panorama de los desafíos de las encuestas de hogares en América Latina

Pedro Silva (ENCE y SCIENCE, Brazil)

La era de los datos

Vivimos en una era en la que la disponibilidad y el acceso a **datos** no tienen precedentes en la historia de la humanidad: **> 400 Millones TB todos los días!**



Estadísticas públicas

Cada vez hay más demanda por datos y estadísticas

más oportunos,

más detallados,

sobre nuevos temas,

de calidad;

para respaldar la toma de decisiones basada en evidencias.

La *datificación* de la sociedad está aumentando la presión sobre las agencias estadísticas, aunque los recursos no aumentan en proporción a la demanda.

Estadísticas públicas

Los datos pueden obtenerse de **fuentes tradicionales** (censos, encuestas por muestreo, registros administrativos) o de **nuevas fuentes orgánicas** (*big data*).

Por décadas, las agencias estadísticas han utilizado las **encuestas probabilísticas** como su herramienta preferida para satisfacer las necesidades de información.

En los últimos años, ha habido un viento de cambio y **nuevas fuentes de datos** están siendo exploradas cada vez más.

Desafíos

Cinco factores clave están detrás de esta tendencia:

- la disminución de las tasas de respuesta en las encuestas probabilísticas;
- el alto costo de la recopilación de datos;
- el aumento de la carga sobre los encuestados;
- el deseo de acceso a estadísticas “en tiempo real”; y
- la proliferación de fuentes de datos “no probabilísticas”.

Beaumont (2020)

Perspectivas

Sin embargo, grande parte de las respuestas a las preguntas de interés siguen proviniendo de las encuestas por muestreo, y en particular, las **encuestas de hogares**.

Cuando los datos provienen de fuentes ‘no planificadas’, hay desafíos para la calidad de la inferencia que pueden proporcionar.

Buenas encuestas son importantes para compensar las deficiencias de las fuentes no probabilísticas, proveyendo ‘muestras de referencia’ para aplicar métodos de análisis que reduzcan los sesgos de cobertura o respuesta.

Panorama

A nivel internacional, no hay ejemplos de países que hayan suprimido las encuestas de hogares por muestreo probabilístico.

Pero hay ejemplos de nuevos usos de los datos de estas encuestas para aplicaciones como:

- Acumulación de datos a lo largo del tiempo;
- Estimación para pequeños dominios;
- Enlaces y la integración con datos de otras fuentes;
- Ponderación de datos de fuentes no probabilísticas.

Desafíos y oportunidades

1. Acumulación de datos a lo largo del tiempo;
2. Estimación para pequeños dominios;
3. Enlaces y la integración con datos de otras fuentes;
4. Ponderación de datos de fuentes no probabilísticas;
5. Aumentar la eficiencia de las muestras y las operaciones;
6. Integrar muestras y registros;
7. Ampliar el uso de herramientas SIG para respaldar la planificación y optimizar las operaciones.

Acumulación de datos a lo largo del tiempo

El ejemplo más destacado de la aplicación de esta idea es, sin duda, la *American Community Survey* (ACS).

Esta encuesta, que comenzó en 2005, sustituyó al cuestionario largo que se aplicaba en los censos de población de Estados Unidos.

Produce resultados **anuales** y **quinquenales** mediante la acumulación de muestras mensuales.

<https://www.census.gov/programs-surveys/acs>

https://www.census.gov/content/dam/Census/programs-surveys/acs/about/2019_Spanish_ACS_Information_Guide.pdf



Acumulación de datos a lo largo del tiempo

Un área que recibe poca atención es la modelización de las series de estimaciones resultantes de encuestas por muestreo repetidas (como las de fuerza laboral).

Estos enfoques ya se utilizan ampliamente en otros países:

- Estados Unidos ([CPS State monthly estimates](#))
- Australia, Holanda, etc.

<https://www.inegi.org.mx/eventos/2024/iaos-isi/doc/19.pdf>

No requiere cambiar las encuestas, pero permite producir nuevos resultados, ampliando el valor de los datos.

→ Conferencia de Denise Silva mañana.

Estimación para dominios pequeños

Vale la pena señalar:

- Esfuerzos de la CEPAL para ampliar el uso de métodos modernos de estimación para dominios pequeños;
<https://statistics.cepal.org/portal/sae/index.html?lang=en>
- Esfuerzos del Banco Mundial para proporcionar métodos y software para la estimación de la pobreza en áreas pequeñas;
<https://pipmaps.worldbank.org/en/data/datatopics/poverty-portal/sae>
- Y del UNSD con el '*toolkit on using SAE for SDGs*':
<https://unstats.un.org/wiki/display/SAE4SDG/SAE4SDG>

Enlaces y la integración con datos de otras fuentes

En el Seminario promovido por CEPAL y INEGI sobre

Integración de Fuentes de Datos y Estadísticas Oficiales

realizado en México (2024) conocimos experiencias de varios países con el tema.

Sin embargo, queda mucho por hacer y muchas otras oportunidades.

→ Conferencia de Denise Silva mañana.

Ponderación de datos de fuentes no probabilísticas

Las muestras no probabilísticas se han utilizado en contextos donde las muestras tradicionales no son viables.

Durante la pandemia de COVID-19, la recopilación de datos cara a cara se volvió poco práctica o imposible.

En Brasil, NIC.br utilizó encuestas a través de paneles web para producir datos sobre los usuarios de Internet.

Los datos producidos presentan un importante sesgo de selección, que fue posible compensar parcialmente mediante el uso de muestras probabilísticas de referencia.

<https://cetic.br/es/pesquisa/tic-covid-19/>

Ponderación de datos de fuentes no probabilísticas

La ponderación de muestras no probabilísticas puede ayudar a compensar el sesgo de selección.

Sin embargo, existen riesgos importantes derivados de los problemas de cobertura.

Investigaciones recientes sobre el tema han generado nuevos métodos, siempre basados en modelos.

Es importante evaluar si se cumplen las condiciones requeridas para el uso de los modelos considerados.

Aumentar la eficiencia de las muestras y las operaciones

El avance de la investigación, la metodología y la tecnología ha permitido aumentar la eficiencia en encuestas de hogares:

- En el diseño de las muestras;
- En la recopilación de los datos;
- En la ponderación de las muestras.

Ilustraré estas ideas con ejemplos de investigaciones realizadas en Brasil.

Formación y estratificación de Unidades Primarias de Muestreo

La muestra maestra de las encuestas por muestreo de hogares del IBGE ha utilizado, desde 2012:

- Métodos de optimización para la formación de UPMs que tengan tamaños mínimos suficientes para sostener la investigación durante un período mínimo de tiempo.
- Métodos óptimos de estratificación para agrupar las UPM en estratos contiguos, minimizando la varianza de las estimaciones del indicador principal (tasa de desempleo).

[Brito \(2024\)](#)

Métodos de Otimização Aplicados em Problemas de Estratificação Estatística

José André de Moura Brito – ENCE/IBGE

Trabalho financiado parcialmente com recursos do Projeto Universal 405044/2021-6.

Distribución de la muestra para su recogida

Una vez seleccionada la muestra de UPMs, es necesario distribuirla entre las unidades de recolección.

El método utilizado hasta ahora se basa en la distribución según las áreas de jurisdicción de las unidades de recolección.

Un trabajo reciente con métodos de optimización muestra que es posible ahorrar un 15% redistribuyendo la asignación de UPMs por unidades de recolección.

Leoni (2024) - <https://github.com/leoniedu/orce>

Um Modelo para Otimizar a Rede de Coleta do IBGE

Eduardo Leoni

eduardo.leoni@ibge.gov.br

IBGE - Superintendência Estadual da Bahia

	Jurisdição	Ótimo Ex Rem	Ótimo c Rem	Projeção	Realizado
N Agencias	47	33	22	41	41
Entrevistadores	96	69	46	75	77
Custo Treinamento	194,437	135,236	87,356	153,279	149,717
Custo Combustivel	98,173	108,648	107,566	103,988	79,928
Custo Diarias	115,575	115,240	212,892	128,472	148,424
Custo Deslocamento	213,748	223,888	320,458	232,460	228,352
Custo Excluindo Remuneracao	408,185	359,124	407,814	385,739	378,069
Custo Incluindo Remuneracao	907,385	717,924	647,014	775,739	778,469
Custo Incluindo Remuneracao Rel	0	-21	-29	-15	-14

Obs: Estamos incluindo no custo de deslocamento aqui somente a combustível e diárias, ou seja, não há custo por hora de viagem.

Ponderación de muestras

Sesgo de disponibilidad: hay proporcionalmente más mujeres y ancianos en la muestra que en la población.

Desde 2021, la ponderación de la muestra del PNADC se realiza con calibración por márgenes, considerando la distribución por sexo y edad.

Esto permitió reducir el sesgo causado por la distribución desequilibrada de la muestra.

Integrar muestras y registros

En muchas encuestas por muestreo, existen oportunidades de integración con datos de registros administrativos.

Por ejemplo, en Brasil se hicieron experimentos de integración con datos del Impuesto a la Renta de las Personas Físicas para mejorar la estimación de la concentración del ingreso.

Estos mostraron que las encuestas de hogares no capturan bien los ingresos de los más ricos.

[Hecksher \(2019\)](#)

Ampliar el uso de herramientas SIG

Hay muchas oportunidades para integrar herramientas SIG para mejorar y apoyar las encuestas por muestreo de hogares.

En la formación de los registros, el IBGE publicó el Registro Nacional de Direcciones con Fines Estadísticos (CNEFE), que por primera vez incluye las coordenadas de las direcciones.

También dio a conocer las rutas de los empadronadores del Censo 2022.

<https://www.ibge.gov.br/estatisticas/sociais/populacao/38734-cadastro-nacional-de-enderecos-para-fins-estatisticos.html?edicao=40122>

Ampliar el uso de herramientas SIG

En Camboya, la OMS utilizó el SIG para generar el registro de direcciones para la selección de muestras de la Encuesta Mundial de Salud Plus (WHS+).

Muestra inicial de UPMs seleccionada con base en el último Censo de Población.

Registro de direcciones de edificios compilado utilizando imágenes satelitales para las UPMs de la muestra en un ambiente de SIG.

Selección de una muestra de coordenadas de edificios enviada al equipo de campo.

[WHO \(2024\)](#)

En conclusión

Hay muchas otras ideas para mejorar y evolucionar las encuestas de hogares – p. ex. [Carletto et al \(2022\)](#):

- 1) Mejorar la **interoperabilidad** y la integración de las encuestas de hogares;
- 2) Diseñar e implementar **encuestas más inclusivas** y centradas en los encuestados;
- 3) Crear capacidad para encuestas CAPI, telefónicas, web y de modo mixto;
- 4) Sistematizar la recopilación, el almacenamiento y el uso de **paradatos** y metadatos;
- 5) Incorporar el aprendizaje automático y la inteligencia artificial para el control y análisis de la calidad de los datos; e
- 6) Mejorar el acceso, la visibilidad y la difusión de los datos.

En conclusión

La **innovación** será clave para permitir que las agencias de estadísticas tengan éxito.

Y para acelerar la innovación es necesario **establecer alianzas** con:

- Otras agencias de estadísticas;
- Otros proveedores de datos;
- Academia.

También será necesario **capacitar** cada vez más a sus equipos técnicos.

Referencias

Beaumont, J.-F. (2020). Are probability surveys bound to disappear for the production of official statistics? *Survey Methodology*, v. 46, n. 1, p. 1-28.

Brito, J.A.M. (2024). *Métodos de Otimização Aplicados em Problemas de Estratificação Estatística*. Rio de Janeiro: Escola Nacional de Ciências Estatísticas.

Carletto, G. et al. (2022). Positioning household surveys for the next decade. *Statistical Journal of the IAOS* 38 (2022) 923–946

Hecksher, M. D. (2019). *Ensaio sobre desigualdade de renda no Brasil*. Rio de Janeiro: Escola Nacional de Ciências Estatísticas, tese de doutorado.

https://ence.ibge.gov.br/images/ence/doc/mestrado/dissertacoes/2019/TeseDoutoradoENCE_MarcosDantasHecksher.pdf

Leoni, E. (2024). *Um Modelo para Otimizar a Rede de Coleta do IBGE*. Rio de Janeiro: IBGE.

WHO (2024). *Enhancing Health Survey Accuracy and Efficiency: A Case Study of GIS Support to the World Health Survey Plus in Cambodia*.

Gracias