

PONDERACIÓN DE UNA ENCUESTA EN PANEL REALIZADA EN EL ESCENARIO DE PANDEMIA

Marcelo Trindade Pitta

Pedro Luis do Nascimento Silva

Webinar "Datos, innovación y producción estadística durante la
pandemia por COVID-19" | 15 Octubre 2020



Organização
das Nações Unidas
para a Educação,
a Ciência e a Cultura

cetic.br

Centro Regional de Estudos
para o Desenvolvimento da
Sociedade da Informação
sob os auspícios da UNESCO

nic.br

Núcleo de Informação
e Coordenação do
Ponto BR

cgi.br

Comitê Gestor da
Internet no Brasil

PRONÓSTICO PARA LAS ENCUESTAS TRADICIONALES

METODOLOGIA Y DESAFIOS



Encuestas tradicionales para estadísticas públicas / oficiales:

- › Son basadas en muestreos probabilísticos seleccionados en registros completos;
- › Recolectadas por medios diversificados: CAPI, CATI y CAWI.

Se ha observado un aumento de las tasas de no respuesta; y una mayor demanda de datos más oportunos y desagregados.

Al mismo tiempo, se cuenta con recursos decrecientes.

Por otra parte....

METODOLOGIAS ALTERNATIVAS

ENCUESTAS NON PROBABLÍSTICAS



Encuestas no probabilísticas:

- › Más baratas de recolectar;
- › Se pueden recolectar muestras más grandes en menos tiempo;
- › Puede introducir algunos tipos de sesgos.

Enfoque para superar las dificultades:

- › Encuesta muestral probabilística como referencia para ponderar la muestra no probabilística;
- › Mitigar el sesgo de las estimaciones: depende del modelo, lo que agrega complejidad a la estimación de la encuesta.



ESCENA ACTUAL

LA PANDEMIA Y LAS ENCUESTAS CARA A CARA



La pandemia de COVID-19 hizo que las tradicionales entrevistas cara a cara fueran imposibles para la mayoría de las encuestas.

Debieron desarrollarse métodos alternativos de recolección de datos para las encuestas de población y de otros tipos.

Los registros incompletos o parciales se han convertido en una barrera fundamental.

RECOLECCIÓN DE DATOS

ENCUESTA NON PROBABILÍSTICA



Encuesta Panel TIC COVID-19

- › **Población objetivo:** usuarios de Internet mayores de 16 años en Brasil.
- › **Estratos:** sexo (2), nivel de educación (3), región (5), grupo de edad (5) y nivel socioeconómico - NSE (4).

Registros:

- › Panel web de individuos mantenido por empresas de investigación de mercado.
- › Listas telefónicas para llegar a una población más amplia (clases sociales menos favorecidas y con menor nivel educativo).

RECOLECCIÓN DE DATOS

ENCUESTA NON PROBABILÍSTICA



Encuesta Panel TIC COVID-19

› **Diseño muestral:** muestra de cuotas según las variables: sexo, nivel de educación, región, grupo de edad y nivel socioeconómico - NSE.

Modalidad	Contactos	Respondientes
Web	18,734	2,539
Teléfono	72,794	88
Total	91,528	2,627

RECOLECCIÓN DE DATOS

ENCUESTA PROBABILÍSTICA



Encuesta TIC Hogares 2019: Encuesta de referencia

- › **Población objetivo:** Hogares brasileños y todas las personas mayores de 10 años que residen en hogares brasileños privados permanentes.
- › **Registro:** IBGE 2010 base de datos de distritos censales.
- › **Tamaño de muestra:** 30.000 entrevistas (hogares e individuos), aproximadamente.
- › **Diseño muestral:** muestreo estratificado de múltiples etapas de hogares y residentes.



METODOLOGIA PONDERACIÓN

CONDICIONES



Soporte común:

El uso de la muestra probabilística para ponderar la encuesta de panel requiere que las dos muestras:

- › Se realicen para el mismo período de referencia;
- › Recopilen una variedad de variables de la misma manera (preguntas superpuestas);
- › Abordar la misma población objetivo;
- › Que no haya intersección de encuestados.

METODOLOGIA PONDERACIÓN

CONDICIONES



Soporte común:

El uso de la muestra probabilística para ponderar la encuesta de panel requiere que las dos muestras:

- ?** › Se realicen para el mismo período de referencia;
- OK** › Recopilen una variedad de variables de la misma manera (preguntas superpuestas);
- ?** › Abordan la misma población objetivo;
- OK** › No haya intersección de encuestados.

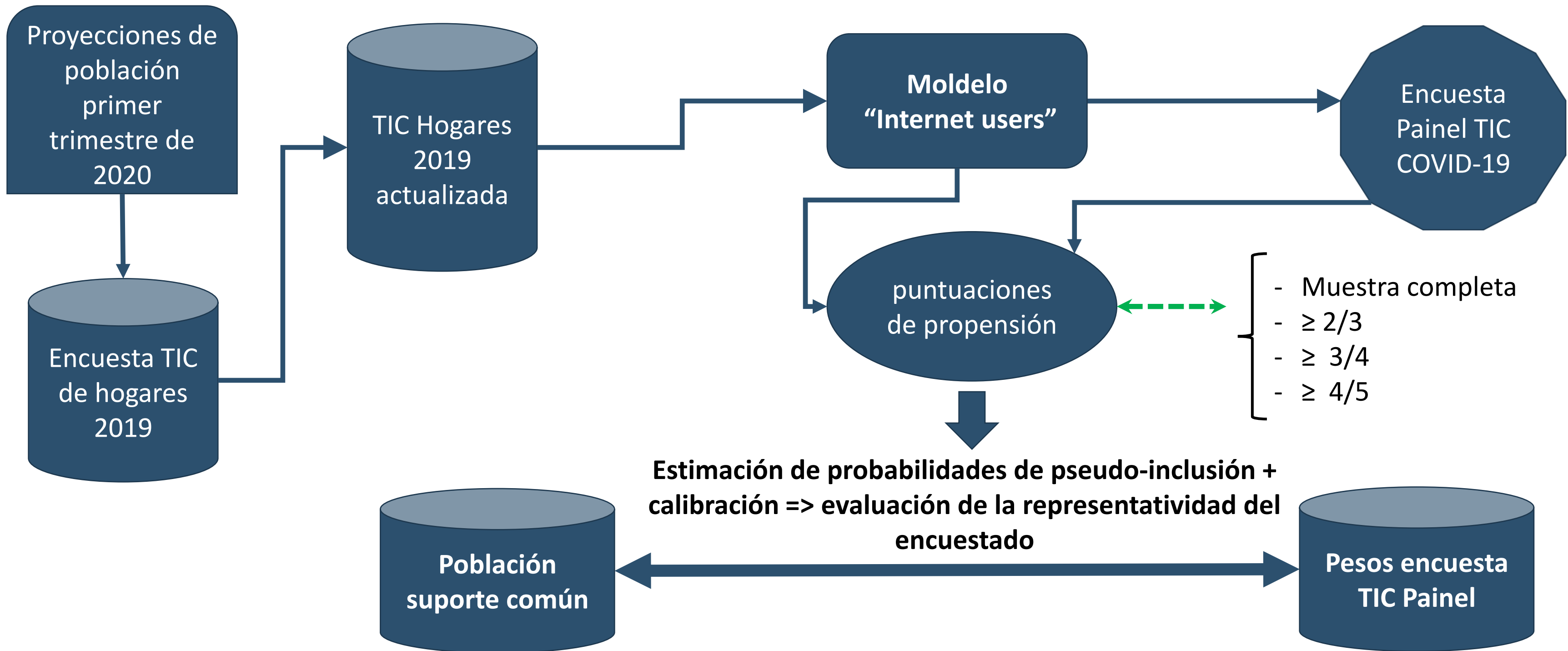
METODOLOGÍA

ENFOQUE UTILIZADO

- › Estimación del tamaño de la población objetivo (usuarios de Internet mayores de 16 años) utilizando los datos recogidos por la Encuesta TIC Hogares, 2019.
- › Estimación de probabilidades de pseudo-inclusión para las unidades muestrales no probabilísticas, mediante un modelo de regresión logística, y utilización de sus recíprocos como ponderaciones.
- › Evaluar e identificar la población representada por los encuestados de la encuesta del panel web, entre los de la población objetivo.
- › Estimar la varianza usando bootstrap.

METODOLOGIA

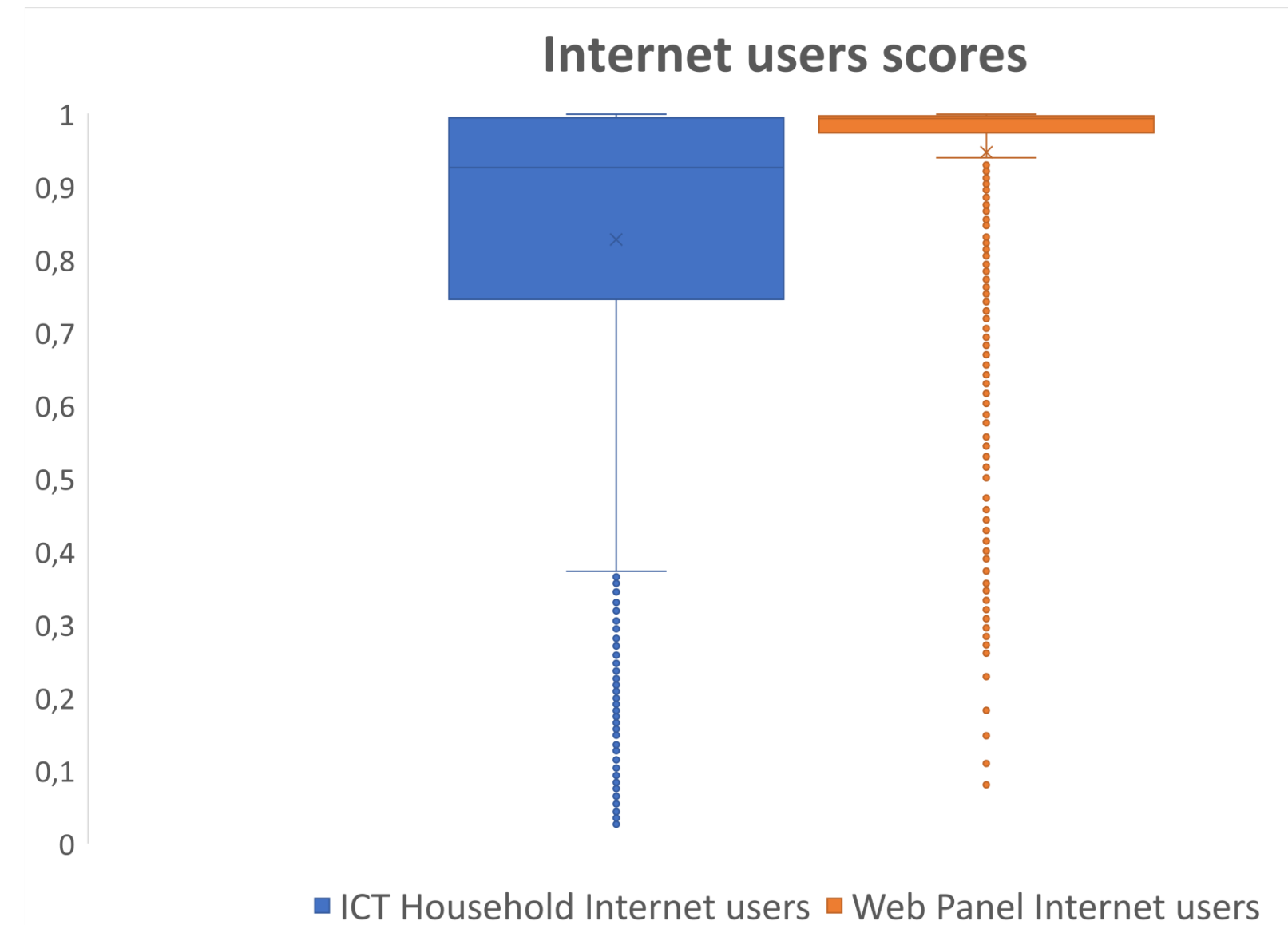
ENFOQUE UTILIZADO



RESULTADOS

PAINEL TIC COVID-19

- › Estadísticas del modelo de usuarios de Internet: $R^2 = 0,431$; Individuos correctamente clasificados = 83%.
- › Comparación de puntuaciones entre la Encuesta de Hogares de TIC y la Encuesta de Panel.



RESULTADOS

PAINEL TIC COVID-19

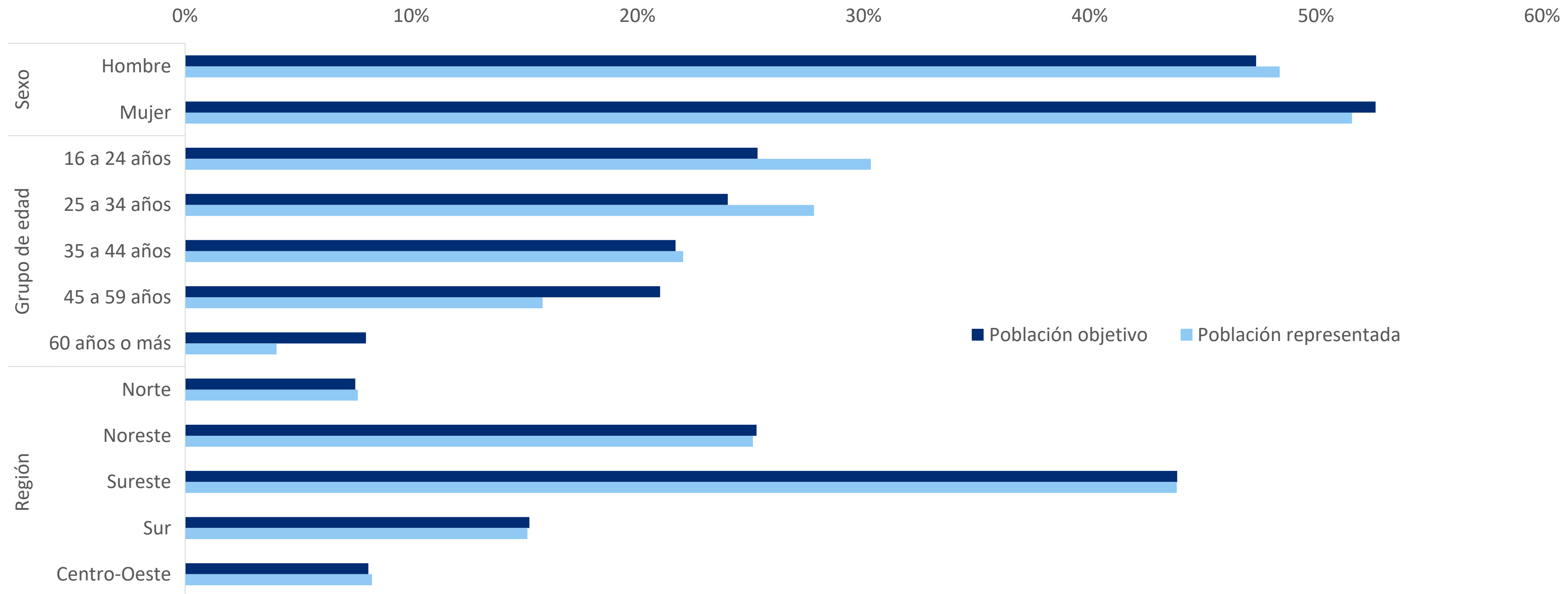
- › Estimación y evaluación de la representatividad de las probabilidades de pseudo-inclusión

Población de muestra de probabilidad utilizada para la calibración

<i>Estadísticas del factor de calibración de peso del Panel</i>	Min.	1ºQu.	Mediana	Media	3ºQu.	Max.
Muestra completa	0.120	0.283	0.369	0.581	0.630	5.840
Muestra con puntuaciones de propensión $\geq 2/3$	0.078	0.484	0.637	0.988	1.038	12.518
Muestra con puntuaciones de propensión $\geq 3/4$	0.063	0.623	0.767	0.918	0.930	4.380
Muestra con puntuaciones de propensión $\geq 4/5$	0.087	0.447	0.554	0.729	0.791	5.604

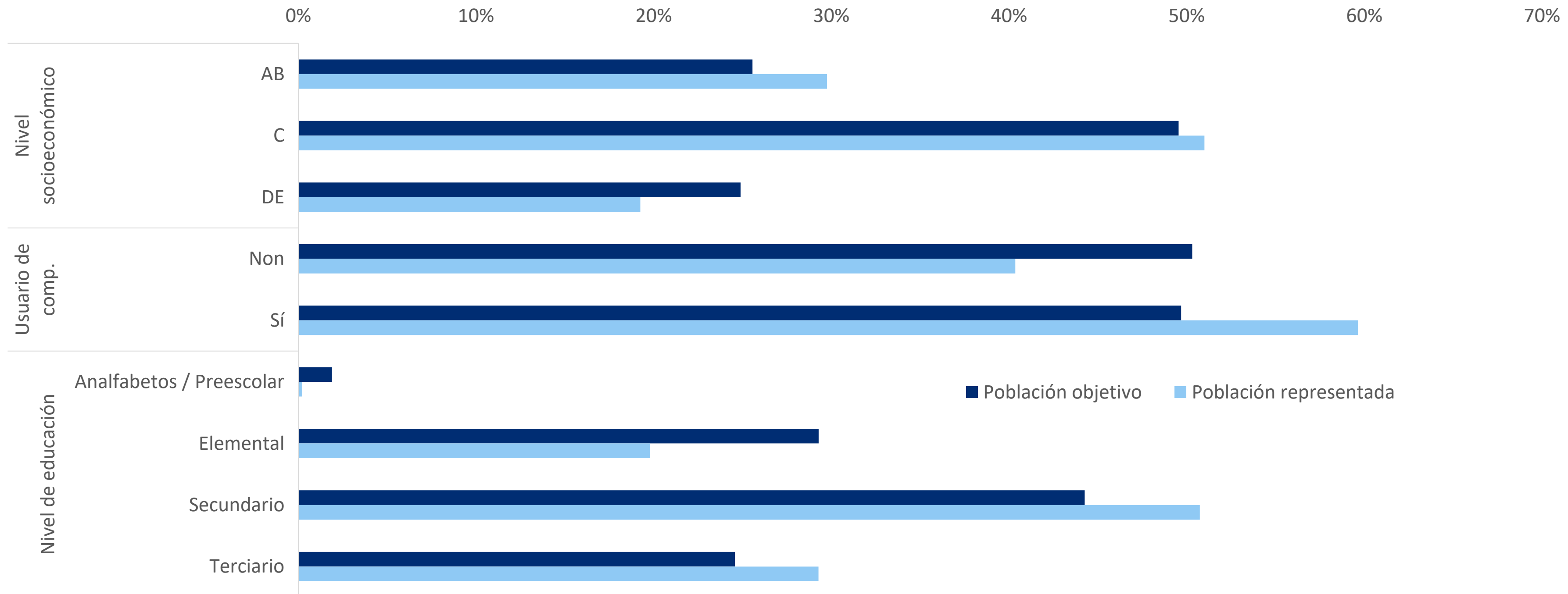
RESULTADOS: Población representada (101M) y objetivo (121M)

PAINEL TIC COVID-19



RESULTADOS: Población representada (101M) y objetivo (121M)

PAINEL TIC COVID-19



RESULTADOS: EVALUACIÓN DE LA EXPERIENCIA

PAINEL TIC COVID-19

Ventajas

- › Los datos se recogieron evitando entrevistas cara a cara.
- › Toda la encuesta, desde la planificación hasta la publicación de los datos, tardó menos de dos meses en completarse.
- › El costo de la recopilación de datos fue mucho más bajo que el de una encuesta tradicional cara a cara.

RESULTADOS: EVALUACIÓN DE EXPERIENCIA

PAINEL TIC COVID-19

Desventajas

- › El reclutamiento de marcos de paneles no está destinado a dirigirse a una población específica.
- › Sigue habiendo problemas de cobertura, a pesar de utilizar una muestra probabilística como referencia.
- › El enfoque depende del modelo; es posible que no siempre haya buenos modelos disponibles.
- › La explicación de la metodología y su difusión es compleja.

CONCLUSIONES

PAINEL TIC COVID-19

- › Está aumentando el uso del muestreo no probabilístico.
- › El uso de una encuesta de muestra probabilística de referencia para ponderar la muestra no probabilística es un intento de producir estimaciones más confiables.
- › Las encuestas por muestreo probabilístico seguirán siendo necesarias, si han de servir como muestras de referencia para ponderar encuestas no probabilísticas más frecuentes.
- › Aún se están desarrollando enfoques para ponderar las muestras no probabilísticas.

¡Muchas Gracias!

Encuesta Painel TIC COVID-19: <https://cetic.br/pt/tics/tic-covid-19/painel-covid-19/1-edicao/>

www.cetic.br
marcelopitta@nic.br
pedronsilva@gmail.com



Organização
das Nações Unidas
para a Educação,
a Ciência e a Cultura

cetic.br **nic.br** **cgi.br**

Centro Regional de Estudos
para o Desenvolvimento da
Sociedade da Informação
sob os auspícios da UNESCO

Núcleo de Informação
e Coordenação do
Ponto BR

Comitê Gestor da
Internet no Brasil

Referencias



- › Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Gile, K. J. and Tourangeau, R. 2013. *Summary Report of the AAPOR Task Force on Non-Probability Sampling*. *Journal of Survey Statistics and Methodology* 1 (2): 90–105. <https://doi.org/10.1093/jssam/smt008>.
- › Coelho, I. B., Pitta, M. T. and Silva, P. L. d. N. 2020. *Estimating State Level Indicators from ICT Household Surveys in Brazil*. *Statistical Journal of the IAOS* 36 (2): 495–508. <https://doi.org/10.3233/SJI-190511>.
- › Dever, J. A. 2018. *Combining Probability and Nonprobability Samples to Form Efficient Hybrid Estimates: An Evaluation of the Common Support Assumption*. In 2018 Federal Committee on Statistical Methodology (FCSM) Research Conference, 15.
- › Elliott, M. R. 2009. *Combining Data from Probability and Non-Probability Samples Using Pseudo-Weights*. *Survey Practice* 2 (6): 1–7. <https://doi.org/10.29115/sp-2009-0025>.

Referencias



- › Elliott, M. R., and Valliant, R. 2017. *Inference for Nonprobability Samples*. *Statistical Science* 32 (2): 249–64. <https://doi.org/10.1214/16-STS598>.
- › ITU, International Telecommunication Union. 2014. *Manual for Measuring ICT Access and Use by Households and Individuals*. https://www.itu.int/dms_pub/itu-d/opb/ind/D-IND-ITCMEAS-2014-PDF-E.pdf.
- › Little, R. J. A., and Rubin, D. B. 2002. *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics.
- › Valliant, R. 2019. *Comparing Alternatives for Estimation from Nonprobability Samples*. *Journal of Survey Statistics and Methodology*, <https://doi.org/10.1093/jssam/smz003>.
- › Valliant, R., and Dever, J. A. 2011. *Estimating Propensity Adjustments for Volunteer Web Surveys*. *Sociological Methods and Research*. Vol. 40. <https://doi.org/10.1177/0049124110392533>.