

Segunda ronda encuesta coronavirus BID-Cornell y panel de datos:

Nota metodológica

Nicolas Bottan
Bridget Hoffmann
Diego A. Vera-Cossio

Departamento de Investigación
y Economista Jefe

NOTA TÉCNICA N°
IDB-TN-2330

Segunda ronda encuesta coronavirus BID-Cornell y panel de datos:

Nota metodológica

Nicolas Bottan*

Bridget Hoffmann**

Diego A. Vera-Cossio**

* Universidad de Cornell

** Banco Interamericano de Desarrollo

Catalogación en la fuente proporcionada por la
Biblioteca Felipe Herrera del
Banco Interamericano de Desarrollo

Bottan, Nicolas.

Segunda ronda encuesta coronavirus BID-Cornell y panel de datos: nota
metodológica / Nicolas Bottan, Bridget Hoffmann, Diego A. Vera-Cossio.

p. cm. — (Nota técnica del BID ; 2330)

Incluye referencias bibliográficas

1. Coronavirus infections-Social aspects-Latin America. 2. Coronavirus infections-
Social aspects-Caribbean Area. 3. Household surveys-Latin America. 4. Household
surveys-Caribbean Area. I. Hoffmann, Bridget. II. Vera-Cossio, Diego A. III. Banco
Interamericano de Desarrollo. Departamento de Investigación y Economista Jefe. IV.
Título. V. Serie.

IDB-TN-2330

<http://www.iadb.org>

Copyright © 2021 Banco Interamericano de Desarrollo. Esta obra se encuentra sujeta a una licencia Creative Commons IGO 3.0 Reconocimiento-NoComercial-SinObrasDerivadas (CC-IGO 3.0 BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) y puede ser reproducida para cualquier uso no-comercial otorgando el reconocimiento respectivo al BID. No se permiten obras derivadas.

Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la CNUDMI (UNCITRAL). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID, no están autorizados por esta licencia CC-IGO y requieren de un acuerdo de licencia adicional.

Note que el enlace URL incluye términos y condiciones adicionales de esta licencia.

Las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.



Resumen¹

La segunda ronda de la Encuesta BID/Cornell Coronavirus hace un seguimiento a la encuesta BID/Cornell Coronavirus de la primera ronda en donde se recolectó información relacionada a los impactos de la pandemia del Covid-19 en los mercados laborales, seguridad alimentaria, distanciamiento social, y conocimiento acerca de los síntomas y formas de contagio del Coronavirus. El objetivo principal de la segunda ronda fue hacer un seguimiento a los hogares participantes de la primera ronda, recolectando información más detallada de ingresos y activos de los hogares, así como también datos de beneficiarios de programas sociales existentes antes de la pandemia, inclusión financiera, preferencias políticas y comportamientos relacionados con el Covid-19 en general. El procedimiento de recolección de datos de la segunda ronda, así como su divulgación, se llevaron a cabo de la misma forma que para el caso de la primera ronda. Los datos de las dos rondas de la encuesta se unieron para crear un panel de datos a nivel de hogar. Al realizar un ejercicio de correlación entre participación en la segunda ronda y características demográficas, encontramos que ser mujer aumentó la probabilidad de participar en la segunda ronda de la encuesta. Así mismo, personas con un nivel de educación más alto y con mayores ingresos fueron más propensas a participar en la segunda ronda. Acorde con estos resultados, encontramos también una correlación positiva entre tener 40 años en adelante y participar en la encuesta de seguimiento. Los datos del panel de datos son útiles para analizar efectos de la pandemia en situaciones y preferencias más detalladas de los hogares dentro de un país y entre países. Se incluyen ponderadores que corrigen diferencias en el tamaño de la muestra entre países y proveen una mayor ponderación a las observaciones de países con mayor población.

Códigos JEL: D31, H55, I14, I32, I38, J01

Palabras clave: encuesta en línea, pandemia del Coronavirus, salud pública, mercado laboral, desigualdad.

¹Nos gustaría agradecer a Sebastián Espinoza, María Paula Medina y David Vargas por su magnífico trabajo como asistentes de investigación. Nos gustaría agradecer a Julián Cristia por su apoyo y asesoramiento en el proyecto. También queremos agradecer a Tom Sarrazin, Sebastián Oliva y Pablo Bachelet por su tremendo apoyo en la disseminación de la encuesta. También nos gustaría agradecer el apoyo de las oficinas de países del IDB, representantes de países y economistas de países por su invaluable trabajo en la personalización de los cuestionarios para cada país. En particular, queremos agradecer a Marta Ruiz Aranz y Diether Beuermann por su apoyo en el lanzamiento de las encuestas en América Central y el Caribe. Quisiéramos agradecer a Ana María Ibáñez, Eric Parrado, Michael Lovenheim, y el equipo de investigadores en el Departamento de Investigación del BID por su valioso apoyo y retroalimentación en diferentes partes del proyecto. **Financiamiento:** Este proyecto fue financiado por los fondos de investigación Coronavirus del BID (RG-E1700-Encuesta Coronavirus) y por el Cornell Atkinson Center for Sustainability. La licencia de Qualtrics y la aprobación IRB son de la Universidad de Cornell.

1 Segunda Ronda

La segunda ronda de la Encuesta sobre Coronavirus de BID-Cornell se realizó en línea entre diciembre del 2020 y febrero del 2021. Al igual que la encuesta de la primera ronda, los módulos se estandarizaron entre países para poder agrupar los datos. El objetivo de la encuesta era recolectar datos a nivel de hogar sobre mercados laborales, ingresos, activos, programas sociales, inclusión financiera, escasez de comida, preferencias políticas y comportamientos relacionados con el COVID-19. Los países participantes de la segunda ronda fueron los siguientes: Bolivia, Chile, Colombia, Costa Rica, Ecuador, El Salvador, México, Panamá, Perú, República Dominicana y Uruguay. El universo de participantes para la encuesta de seguimiento fueron los participantes de la primera ronda que compartieron su correo electrónico (aproximadamente 116,000 participantes). De este universo de participantes, el número total de personas que respondieron la segunda ronda de la encuesta, controlando por inconsistencias en las respuestas², fue de 12,362.

A menos que se indique lo contrario en este documento, en la segunda ronda se siguieron los mismos procedimientos de limpieza de datos, creación de ponderaciones de las observaciones, etc., que en la primera ronda. La documentación detallada de la encuesta de la primera ronda puede encontrarse [aquí](#).

Table 1: Fecha de Lanzamiento y Número de Observaciones por País

País	Fecha de Lanzamiento	No. de observaciones
Bolivia	12/22/20	591
Chile	12/22/20	2,323
Colombia	12/22/20	1,948
Costa Rica	12/22/20	704
República Dominicana	12/22/20	410
Ecuador	12/22/20	695
El Salvador	12/22/20	643
México	12/22/20	1,258
Panamá	12/22/20	1,065
Perú	12/22/20	1,155
Uruguay	12/22/20	1,570

Fecha de lanzamiento es la fecha en la que las publicaciones en redes sociales empezaron. La fecha en la que las encuestas fueron lanzadas en cada país fue determinada por procesos y aprobaciones burocráticas en cada país.

² Siguiendo los mismos criterios que se usaron para la primera ronda, se eliminaron las encuestas cuya duración fue extremadamente corta (menos de 9 minutos) o extremadamente larga (más de 30 horas). También, se restringieron las respuestas válidas a las personas que respondieron correctamente la pregunta de chequeo de atención y a las personas que respondieron la encuesta en su totalidad.

2 Distribución de la encuesta

La encuesta de la segunda ronda se distribuyó por medio de correo electrónico. Se le envió un enlace individual a los participantes de la primera ronda que habían indicado previamente estar dispuestos a ser contactados para una encuesta de seguimiento y que también proporcionaron su dirección de correo electrónico. Se enviaron también recordatorios adicionales a los participantes por correo electrónico.

El mensaje que se incluyó en los correos electrónicos se presenta en la Figura 1.

Figura 1. Mensaje Correo Electrónico



3 Generación de nuevas variables

- Debido a la presencia de múltiples errores de tipografía en preguntas de la segunda ronda relacionadas con el ingreso del hogar, se construyeron nuevas variables de ingreso corregidas siguiendo los siguientes criterios:
 1. Los ingresos que fueron ingresados en letras fueron cambiados a números.
 2. Los ingresos con respuestas como “nada”, “cero”, fueron reemplazados por un ingreso de cero.
 3. Los ingresos con respuestas como “Más de 100 mil”, “Menos de 100 mil”, fueron transformados a valores “missing” ya que su precisión es muy poca.
 4. Los ingresos con respuestas como “500 a 600 mil pesos”, fueron transformados a un promedio entre los dos valores registrados.
 5. Finalmente, los ingresos con respuestas inconsistentes fueron transformados a valores “missing”.

Posteriormente, se construyeron variables categóricas de ingreso con base en rangos del salario mínimo de cada país. Los rangos utilizados en las variables de ingreso fueron los siguientes: menos de 1 salario mínimo; entre 1 y 2 salarios mínimos; entre 2 y 3 salarios mínimos; entre 3 y 4 salarios mínimos; entre 4 y 6 salarios mínimos; entre 6 y 8 salarios mínimos; entre 8 y 11 salarios mínimos; más de 11 salarios mínimos.

- Para hacer una distinción entre las variables que se preguntaron en las dos rondas de la encuesta (ingreso, género y edad), se añadió un sufijo a las variables. Si el sufijo es `_r1` significa que la variable pertenece a la primera ronda, si el sufijo es `_r2` significa que la variable pertenece a la segunda ronda.

4 Panel de Datos

Los datos de la primera y segunda ronda se unieron para crear un panel de datos a nivel de hogar. A grandes rasgos, de las 230,509 personas que participaron en la primera ronda, 11,492 participaron también en la segunda ronda. Para controlar por posibles desbalances entre los datos suministrados por el panel de datos y los datos de la población general de cada país, construimos variables de ponderación para cada observación.

4.1 Validación

4.1.1 Estimación de pesos

Al igual que para el caso de la primera ronda, modelamos la probabilidad de ser parte de la encuesta representativa nacionalmente para cada encuestado i en el país c . Utilizamos características que no varían en el tiempo como género, edad y educación del encuestado, indicadores de presencia en el hogar de niños de 5 años o menos y de adultos mayores (60 años o más), así como el número total de miembros del hogar, el número de niños que viven en el hogar y el número de niños matriculados en la escuela.

Realizamos este proceso país por país. Para cada uno, usamos la encuesta nacional representativa más reciente disponible en el repositorio de bases armonizadas del Banco Interamericano de Desarrollo. Estimamos las probabilidades predichas de ser parte de la encuesta nacional representativa con base en un modelo de probabilidad lineal. Escogimos un modelo diferente al utilizado en el caso de la primera ronda por motivos de no-convergencia en la mayoría de países al utilizar un modelo ajustado a una función logística (logit). Para evitar probabilidades negativas y/o probabilidades mayores que uno reajustamos las probabilidades predichas con base en la siguiente fórmula:

$$\text{Probabilidad predicha ajustada} = \frac{\text{prob. predicha inicial} - \min(\text{prob. predicha inicial})}{\max(\text{prob. predicha inicial}) - \min(\text{prob. predicha inicial})} \quad (1)$$

Con las probabilidades predichas ajustadas para cada encuestado i en el país c , usamos el inverso de probabilidad de los pesos ($ipw_{i,c} = 1/(1 - \hat{p}_{i,c})$) para que, al menos en términos de características observables, las observaciones del panel de datos sean similares a las encuestas de hogares representativas.

Posteriormente, usamos los propensity scores ($ipw_{i,c}$) como semillas para la calibración de pesos usando un método proporcional de ajuste iterativo (RAKING). Con este método de ponderación, el investigador elige un conjunto de variables cuya distribución poblacional se conoce, y el procedimiento ajusta iterativamente la ponderación de cada caso hasta que la distribución de la muestra se alinea con la población para esas variables (Mercer et al., 2018). Para nuestro caso, trabajamos con un conjunto de variables estándar que difiere en ciertos casos para cada país por motivos de no-convergencia de la estimación de los pesos al utilizar todo el conjunto de variables. A continuación, listamos el conjunto de variables utilizado para cada país:

1. Uruguay, República Dominicana, Chile: proporción de mujeres, categoría de edad (< 40, 40-60, > 60 años), categoría de educación (primaria, secundaria y universitaria), categoría de ingreso (basado en 8 niveles de ingreso con base en el salario mínimo de cada país).
2. Perú, Colombia, Costa Rica, México: proporción de mujeres, categoría de edad, categoría de ingreso.
3. Bolivia: proporción de mujeres, categoría de edad, categoría de educación.
4. Ecuador: proporción de mujeres, categoría de educación, categorías del número de niños que viven en el hogar (0, de 1 a 3, más de 3).
5. El Salvador: proporción de mujeres, categoría de edad, categorías del número de niños que viven en el hogar.
6. Panamá: proporción de mujeres, categoría de edad.

Después de realizar el proceso de RAKING, el número de observaciones de la encuesta en línea de cada país coincide con el número total de observaciones de la encuesta nacional representativa de cada país. Para prevenir que algunos países estén sobre-representados y otros sub-representados en la muestra total, re-escalamos los pesos dentro del país por el inverso de la proporción del número reponderado de respuestas por país, relativo a la población total del país ($POB_c / \sum_{i=1}^{N_c} w_{i,c}$) según el censo nacional.

Algunos estudios han demostrado que realizar el proceso de ponderación combinando el método de "propensity score" y el método proporcional de ajuste iterativo, puede ser más eficiente para reducir el sesgo que cualquier otro método aplicado por separado (Dutwin and Buskirk, 2017). El método de "propensity score" no consigue que la muestra coincida exactamente con la población en todas las dimensiones, pero este tipo de modelos pueden detectar relaciones entre las variables de ajuste que el método proporcional de ajuste iterativo no tendría en cuenta. Estimar en una primera etapa el "propensity score" y en una segunda etapa el método de RAKING puede mantener esas relaciones y hacer que la muestra se asemeje completamente con los márgenes de la población (PEW Research Center, 2018).

4.1.2 Número total de observaciones

De acuerdo a los números presentados en la Tabla 2, la suma total de las variables de ponderación en la encuesta en línea y en las encuestas nacionales representativas coinciden para cada país. Con lo anterior, comprobamos que el número total de observaciones en las encuestas nacionales representativas es igual al número total de observaciones de la encuesta en línea al utilizar la variable de ponderación construida previamente.

Tabla 2. Encuesta de Seguimiento en Línea y Datos de Registros Nacionales

País	Encuesta en línea		Encuestas de hogares	Datos del censo
	(1) Variable de ponderación	(2) Variable de ponderación (entre países)	(3) Variable de ponderación	(4) Población total
Bahamas				
Barbados				
Bolivia	7,320,654	11,633,371	7,320,654	11,633,371
Chile	13,413,529	18,946,758	13,413,529	18,946,758
Colombia	32,618,042	50,372,428	32,618,042	50,372,424
Costa Rica	3,710,171	5,111,221	3,710,171	5,111,221
Ecuador	10,975,406	17,510,642	10,975,406	17,510,643
El Salvador	4,557,576	6,581,411	4,557,576	6,581,411
Guyana				
Jamaica				
México	85,763,600	125,016,448	85,763,600	125,016,449
Panamá	2,703,268	4,926,295	2,703,268	4,926,295
Perú	22,339,826	32,824,356	22,339,826	32,824,358
República Dominicana	6,883,808	10,448,499	6,883,808	10,448,499
Suriname				
Trinidad and Tobago				
Uruguay	2,622,211	3,530,912	2,622,211	3,530,912

Las columnas (1) y (2) presentan la suma total de las variables de ponderación construidas para todas las observaciones de la encuesta en línea. La columna (1) muestra la suma de la variable de ponderación sin ser reajustada por la población total. La columna (2) por el contrario, muestra el total de la variable de ponderación reajustada por el número de población total. Como variables de comparación, las columnas (3) y (4) presentan la sumatoria de la variable de ponderación para la encuesta nacional representativa de cada país y la población total según el censo, respectivamente.

4.1.3 Estadísticas descriptivas

Por otro lado, validamos también los pesos en la encuesta en línea comparando las medias de las características demográficas con y sin pesos ponderados en la encuesta en línea, en el panel de datos, y en las encuestas nacionalmente representativas. La Tabla 3 muestra que tanto para la encuesta en línea de la primera ronda como para el panel de datos, los individuos que responden la encuesta son más educados y probables de ser mujeres que los que responden las encuestas de hogares (ver columnas (1), (3), (5) y (7), respectivamente). Las columnas (2), (4) y (6), evidencian que, al utilizar las variables de ponderación recalibradas, las diferencias en las características de los encuestados en las encuestas en línea y las encuestas nacionales representativas, se reducen en la mayoría de las variables demográficas analizadas.

Tabla 3. Diferencias entre Encuestas en Línea y Encuestas de Hogares

Variable	Encuesta en línea						Encuestas de hogares
	R1		R1-email		Panel		Total
	N.P (1)	P (2)	N.P (3)	P (4)	N.P (5)	P (6)	
Tamaño del hogar	4.39	4.45	4.38	4.41	3.94	4.11	3.86
Adulto mayor en el hogar (prop)	0.35	0.34	0.35	0.34	0.39	0.45	0.36
Mujer (prop)	0.72	0.57	0.70	0.55	0.70	0.52	0.52
Educación: Primaria o menor	0.03	0.17	0.03	0.14	0.01	0.06	0.42
Educación: Secundaria	0.24	0.49	0.24	0.49	0.13	0.24	0.41
Educación: Universitaria o mayor	0.73	0.34	0.73	0.36	0.86	0.70	0.16
Edad	39.00	40.79	38.98	40.53	41.66	43.11	43.49
Ingreso: Menos de 1 SM	0.26	0.18	0.26	0.16	0.18	0.17	0.18
Ingreso: 1-2 SM	0.22	0.25	0.22	0.24	0.19	0.23	0.24
Ingreso: 2-3 SM	0.14	0.12	0.14	0.11	0.14	0.17	0.18
Ingreso: 3-4 SM	0.10	0.15	0.11	0.16	0.12	0.12	0.12
Ingreso: 4-6 SM	0.10	0.13	0.10	0.14	0.13	0.15	0.14
Ingreso: 6-8 SM	0.06	0.08	0.06	0.09	0.08	0.06	0.06
Ingreso: 8-11 SM	0.05	0.05	0.05	0.06	0.07	0.06	0.04
Ingreso: 11+ SM	0.06	0.05	0.06	0.05	0.08	0.05	0.04

La tabla presenta promedios de características demográficas usando datos de las dos rondas de la encuesta en línea y las encuestas nacionales representativas. Las columnas (1), (3) y (5) muestran los promedios sin ajustar por la variable de ponderación. Las columnas (2), (4) y (6) muestran los promedios luego de ponderar las observaciones de las encuestas en línea. La columna (7) reporta medias para todas las observaciones disponibles en la encuesta nacional representativa de cada país usando los pesos de muestreo. Las columnas (3) y (4) hacen referencia a los participantes de la primera ronda que compartieron su correo (universo de participantes para la encuesta de seguimiento). SM se refiere al salario mínimo de enero 2020.

4.1.4 Correlación de variables demográficas y participación en la segunda ronda

Realizamos un ejercicio de correlación entre las principales variables demográficas y la decisión de participación en la segunda ronda de la encuesta. La Tabla 4 muestra los resultados de estimar un modelo de probabilidad de lineal para la participación en la segunda ronda de la encuesta. Estimamos esta probabilidad para todos los participantes de la primera ronda (columnas 1 y 2) y para los participantes de la primera ronda que compartieron su correo electrónico (columnas 3 y 4). De acuerdo con estos resultados, ser mujer aumenta la probabilidad de participar en la encuesta en todos los escenarios, así como también lo hace vivir en un hogar con menos integrantes (los hogares con más de 6 integrantes son menos propensos a participar en la segunda ronda relativo a los hogares con menos de 6 integrantes). Relativo a las personas en el grupo de comparación para las categorías de educación (educación primaria o menor), tener un nivel de educación mayor aumenta la probabilidad de participar en la segunda ronda. Así mismo, relativo a las personas en el grupo de comparación para las categorías de ingreso (ingreso en enero 2020 de menos de 1 salario mínimo), tener un ingreso mayor a 1 salario mínimo aumenta la probabilidad de participar en la segunda ronda.

A grandes rasgos, los resultados de la Tabla 4 evidencian que personas con un nivel de educación mayor y con mayores ingresos son más propensas a participar en la encuesta. Acorde con estos resultados, encontramos también una correlación positiva entre tener 40 años en adelante y participar en la segunda ronda.

Tabla 4. Correlación Demográficas y Participación en Segunda Ronda

Variables	(1)	(2)	(3)	(4)
	Prob. segunda ronda			
Tamaño del hogar > 6	-0.013*** (0.001)	-0.007*** (0.001)	-0.024*** (0.003)	-0.014*** (0.003)
Adulto mayor en el hogar	-0.002* (0.001)	0.001 (0.001)	-0.003* (0.002)	0.002 (0.002)
Mujer	0.008*** (0.001)	0.005*** (0.001)	0.023*** (0.002)	0.018*** (0.002)
<i>Educación</i>				
Secundaria	0.011*** (0.003)	0.017*** (0.003)	0.020*** (0.005)	0.032*** (0.005)
Universitaria o mayor	0.025*** (0.003)	0.030*** (0.003)	0.043*** (0.005)	0.056*** (0.005)
<i>Edad</i>				
40 a 59 años	0.015*** (0.001)	0.012*** (0.001)	0.029*** (0.002)	0.023*** (0.002)
60 años o más	0.025*** (0.002)	0.017*** (0.002)	0.055*** (0.004)	0.044*** (0.004)
<i>Ingreso enero 2020</i>				
1-2 SM	0.010*** (0.001)	0.010*** (0.001)	0.019*** (0.002)	0.019*** (0.002)
2-3 SM	0.017*** (0.001)	0.018*** (0.002)	0.034*** (0.003)	0.034*** (0.003)
3-4 SM	0.027*** (0.002)	0.029*** (0.002)	0.052*** (0.003)	0.052*** (0.003)
4-6 SM	0.038*** (0.002)	0.040*** (0.002)	0.070*** (0.003)	0.070*** (0.003)
6-8 SM	0.042*** (0.002)	0.046*** (0.002)	0.077*** (0.004)	0.080*** (0.004)
8-11 SM	0.050*** (0.003)	0.056*** (0.003)	0.091*** (0.005)	0.096*** (0.005)
11+ SM	0.047*** (0.002)	0.056*** (0.002)	0.095*** (0.005)	0.104*** (0.005)
Efectos fijos de país	No	Sí	No	Sí
Muestra	R1	R1	R1-email	R1-email
Observaciones	226,206	226,206	115,462	115,462
R cuadrado	0.011	0.021	0.021	0.036

La tabla muestra los resultados de estimar un modelo de probabilidad lineal para la participación en la segunda ronda de la encuesta. El grupo de comparación para educación es tener un nivel de educación igual a primaria o menor; el grupo de comparación para ingreso es haber tenido un ingreso en enero 2020 igual a menos de 1 salario mínimo; el grupo de comparación para edad es personas entre 18 y 39 años de edad. Las columnas (1) y (2) presentan los resultados para toda la muestra de la primera ronda. Las columnas (3) y (4) muestran los resultados para los participantes de la primera ronda que dieron su correo electrónico. *, ** y *** denotan significancia estadística al nivel del 10%, 5% y 1%, respectivamente. Errores estándar entre paréntesis.

4.2 Uso de pesos para cálculos utilizando el panel de datos

- Para estadísticas a nivel país: usar la variable "*weight_panel*" para corregir por el método de muestreo.
- Para estadísticas entre países: primero obtenga promedios a nivel país usando la variable de ponderación "*weight_panel*" y luego obtenga promedios entre países ponderando por la población de cada país (variable *pop*).

References

Dutwin, D. and Buskirk, T. D. (2017). Apples to Oranges or Gala versus Golden Delicious?: Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples. *Public Opinion Quarterly*, 81(S1):213–239.

Mercer, A., Lau, A., and Kennedy, C. (2018). For weighting online opt-in samples, what matters most? Pew Research Center.