



19/09/2019

No maten al encuestador

TXT [SOL MINOLDO, DIEGO BORN](#) IMG [RO MAGA](#)

¿Qué pasó con el INDEC? ¿Es posible reconstruir los datos que nos faltan?

En los últimos años, puntualmente desde 2007, escuchamos a menudo hablar del **INDEC** (Instituto Nacional de Estadísticas y Censos): que fue intervenido, que tergiversaban los datos, que no era creíble o confiable. Y si algo podemos rescatar de eso es que **de pronto muchísimas personas tomaron conciencia de la existencia y la importancia de las estadísticas públicas**. Pero también nos obligó a enfrentar el hecho de que los datos pueden ser, en sí mismos, un problema. Y que parte de la calidad de los datos tiene que ver con la confianza que generan los procesos con los que son producidos.

Después de cerrar en 2015 el ciclo conocido como ‘la intervención del INDEC’, sobrevive el sabor amargo de varios años en los que los datos quedaron bajo

sospecha. Y son muchas las discusiones que se estancan cuando, para analizar un tema, queremos referirnos a la evidencia y alguien responde ‘¡Ah, pero ese dato no es confiable!’. Frente a eso, ¿qué hacemos?

Los hechos

Cuando se trata de elaborar información estadística, el Estado tiene un rol central. En el caso de Argentina, el INDEC es la institución que encabeza desde 1968 el Sistema Estadístico Nacional. Durante muchos años, el INDEC no dejó de ganar reconocimiento por el alto estándar de sus técnicos y la calidad de su información. Pero en 2007 la reputación del instituto se fue a pique.

Para entender lo que pasó tenemos que retroceder un poquito más, al año 2006. Por ese entonces, los funcionarios políticos de la Secretaría de Comercio Interior no estaban conformes con los datos de inflación publicados por el instituto y les solicitaron información a los técnicos: querían saber de dónde habían sacado los precios, es decir, literalmente las direcciones de los comercios donde habían hecho el relevamiento. Los técnicos entendieron que proporcionar esa información violaba el secreto estadístico y se negaron.

Las asperezas entre técnicos y funcionarios políticos no eran una novedad en la historia del INDEC, pero en esta ocasión el desenlace fue escandaloso. Al ver que la tensión escalaba y las autoridades del INDEC no cedían, se decidió suprimir la discordia procediendo a cambiar a dichas autoridades con el decreto 100/07. Y junto con los cambios de autoridades se produjeron despidos y desplazamientos de los especialistas que confrontaban con sus directivas. Fue tal la falta de legitimidad de estos hechos que, aunque no lo fuera en términos legales, se extendió y sostuvo en el tiempo la idea de que se trataba de una intervención (política) del organismo. El gobierno nacional alegaba que el cambio de autoridades había tenido que ver con una necesidad técnica para implementar un cambio metodológico en la medición de la evolución de precios. Ciertamente estaba previsto avanzar en cambios relevantes, pero que la nueva gestión hubiera ayudado a mejorar el proceso resultó bastante inverosímil. Y, de hecho, en 2010 quedó prácticamente confirmado que los cambios habían deteriorado el funcionamiento del instituto.

Ese año se conoció el informe del Consejo Académico de Evaluación y Seguimiento (CAES) del INDEC, que había convocado el propio gobierno frente a la creciente falta de credibilidad del instituto. El CAES estaba integrado por representantes de varias Universidades Públicas Nacionales que fueron invitados por el Ministerio de Economía y Finanzas Públicas. En su informe final, el Consejo concluyó que los cambios realizados tenían que ver con una discusión institucional más que académica o metodológica, que el instituto había deteriorado la calidad de su personal en términos de capital humano (despidiendo valiosos recursos e incorporando personal con menos formación y trayectoria) y que la información producida a partir de entonces era imprecisa y cuestionable en varios de los indicadores, pero sobre todo en el caso del **Índice de precios**.

Al día de hoy **hay un enorme consenso sobre el hecho de que el índice de precios del INDEC no reflejaba la realidad en aquel entonces**. En definitiva, ni siquiera el propio gobierno parecía dar crédito a los datos sobre precios del instituto, como se desprendía en la negociación salarial de cada año entre 2007 y 2015 (que se establecía en niveles muy superiores a la supuesta inflación informada por el INDEC). Por suerte, el vacío no es absoluto: quedaron los institutos de estadísticas de algunas provincias, que permitieron usar como referencia, por ejemplo, la inflación de San Luis, o un promedio de la inflación en las 7 provincias con estadísticas confiables, o los datos de CABA desde 2012. Además, hubo constante relevamiento por parte de actores privados.

Pero más allá de lo que ocurrió con los precios, las sospechas afectaron al INDEC en su conjunto. Y ahí es donde entra en escena la fuente que particularmente nos preocupa: la **Encuesta Permanente de Hogares (EPH)**.

Las sospechas

La EPH es una fuente de información fundamental para los que trabajamos en ciencias sociales con técnicas cuantitativas que, a diferencia de las técnicas cualitativas, procuran construir indicadores numéricos para observar y analizar la realidad, aspirando a que ese indicador sea representativo de lo que sucede en la sociedad (para lo cual hace falta contar con datos estadísticos). Nos proporciona

información muy completa sobre ingresos y mercado de trabajo. **A diferencia del censo, la EPH es una encuesta que no se realiza a toda la población sino a una muestra;** es decir, a una selección de individuos que se espera que sean representativos de toda la población (en este caso, de toda la población que vive en ciudades). Pero **mientras el censo se realiza sólo cada 10 años, la EPH nos proporciona información de manera más continua (cada tres meses).**

La cuestión es que con la pérdida de credibilidad del INDEC en general, la EPH quedó también bajo sospecha. De hecho, revisando trabajos académicos y de consultoras, y observaciones del INDEC posterior a 2015, se pueden identificar 10 sospechas concretas. Algunas ponen en duda los datos sobre ingresos, que son claves por ejemplo para medir la pobreza, la desigualdad o la evolución de los ingresos en general o en ciertos sectores de la sociedad. Se hipotetizó que los datos fueron manipulados sin más, o que se introdujeron sesgos al hacer cambios metodológicos, o que perdieron representatividad al crecer la falta de declaración de ingresos en esos años.

Otras sospechas tienen que ver con datos de la realidad laboral, suponiendo que se puede haber ‘escondido’ parte del desempleo, identificando a las personas desocupadas como si fueran ocupadas o como ‘inactivas’ (es decir, que no tienen trabajo, pero que tampoco lo están buscando).

Por último, también se generaron sospechas sobre temas bastante técnicos (problemas de cobertura geográfica, inconsistencias muestrales, cambios metodológicos poco transparentes y cambios en las proyecciones de población), que en definitiva implican que los datos, y la evolución que se ve en esos años, no reflejan correctamente la realidad social.

¿Y entonces?

La reconstrucción

Después de 2015, la nueva gestión del INDEC sacó un comunicado recomendando que esa información ni la miremos (literalmente, dijeron que desestimaban su utilización).

Estamos hablando de no poder analizar empíricamente la realidad social de un periodo de nuestra historia reciente y aceptar algo así como un agujero estadístico irrecuperable. ¿No había una opción menos drástica que descartar por completo toda la información producida esos años? ¿Qué pasa si, en vez de tirar la toalla, convertimos las sospechas en hipótesis e intentamos chequearlas? Si pudiéramos establecer para qué lado van las limitaciones de los datos, y su alcance, se volverían utilizables, a conciencia de los márgenes de error y posibles sesgos.

Siguiendo estas ideas, empezamos a trabajar en un proyecto de investigación donde, a grandes rasgos, lo que hicimos fue revisar todas las fuentes estadísticas de carácter público producidas independientemente del INDEC. O sea, no sospechadas. Y en base a los datos que nos daban, sobre cuestiones específicas, compararlos con los datos de la EPH, para ver si se parecían o no (este trabajo culminó con la publicación del libro Claroscuros al que puede accederse gratis [acá](#)).



¿DATOS FALTANTES?

**NO ESTÁN FALTANTES
SÓLO SON RELATIVAMENTE
DESCONOCIDOS**

También miramos los datos de la EPH en el punto de partida (2003-2006) y el punto de llegada (2016-2018), para ver si lo que pasó en medio tenía algún tipo de continuidad. De hecho, nos dimos cuenta de que era muy importante entender lo que pasó con los datos de 2015 a 2016: porque los problemas no se resuelven de un día para otro, entonces corregir sesgos, desprolijidades y limitaciones varias puede ir afectando los datos desde 2016, mostrando cambios que no reflejan

transformaciones de la realidad, sino cambios metodológicos o mejoras en la calidad de los relevamientos, registro y procesamiento de la información.

El objetivo no era analizar qué pasó con el empleo y la pobreza y todas las variables que se pueden mirar en la EPH, sino conocer la solidez y limitaciones de esos datos, para recién después saber hasta dónde se los puede tomar en cuenta y pasar a lo que el experto en desigualdad y pobreza Alberto Minujín señala como lo realmente relevante: discutir, sobre la base de los mismos datos, por qué pasó lo que pasó en la realidad social.

Y la verdad es que **los resultados nos llevaron a concluir que descartar la información de esos años sería, como mínimo, una exageración.** O como lo expresa el investigador Daniel Schteingart, la EPH de esos años “puede usarse para muchas cosas, siendo conscientes de sus defectos y de distintos cambios que fue teniendo a lo largo del tiempo, que dificultan su comparabilidad”. Incluso, encontramos fuertes indicios de que no hubo una ‘manipulación de los datos’ destinada a mejorar los indicadores sociales. Por el contrario, la desprolijidad llevó a que, en algunos casos, los indicadores sociales fueran incluso peor en los números del INDEC que en la realidad.



Como muestra sobra un botón

El dato de ‘ingresos’ es la plata que cada mes entra en la economía del hogar, ya sea en concepto de salarios, jubilaciones, becas, ayudas sociales, cuotas alimenticias, etc. Esa información nos permite, entre otras cosas, medir pobreza, saber el nivel de ingreso de los trabajadores, si ganaron o perdieron en un determinado año con una determinada inflación, y también saber cómo evoluciona la desigualdad. Si quisiéramos manipular el dato de ingresos para que los indicadores sociales nos den mejor, tendríamos que ‘sobrestimar’ los ingresos: o sea, registrar los ingresos en un nivel más alto de lo que son en realidad. Pero los resultados nos dieron, justamente, lo contrario.

El deterioro en el relevamiento de los ingresos existió, pero el sesgo que eso produjo fue de una ‘subrepresentación’ de los ingresos, y no al revés. O sea, **se registraron ingresos más bajos de lo que eran en realidad.**

Ingresos laborales

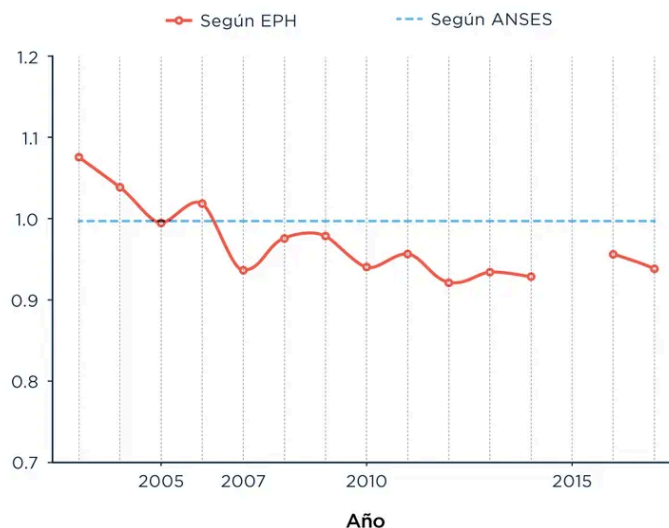


En este gráfico comparamos los datos sobre ingresos laborales en EPH con los de otra fuente, OEDE (Observatorio de Empleo y Dinámica Empresarial). En cada punto de la serie, los valores de OEDE están representados con el valor 1. Los de EPH están representados como la proporción que representan de esa unidad. Para interpretar el gráfico tenemos que pensar que, cuanto más horizontales evolucionen las líneas, va a significar que más estable va a ser la relación entre los datos de la EPH y los de la otra fuente. Pero lo que vemos es que, desde 2007, los ingresos de la EPH son cada vez más bajos que los de la otra fuente. Por eso, representan un porcentaje cada vez menor de los datos en esa fuente. Después de 2015 vemos que hay, abruptamente, una

pequeña mejora en esa relación.

Una aclaración: el OEDE pertenece al (ex) Ministerio (ahora Secretaría) de Trabajo, Empleo y Seguridad Social (MTEySS). Los datos de OEDE usados acá se basan en los que le aporta el Sistema Integrado de Jubilaciones y Pensiones (SIPA).

Ingresos jubilatorios



En este gráfico comparamos los datos de ingresos jubilatorios en EPH con los de ANSES (Administración Nacional de la Seguridad Social). Como en el gráfico de arriba, si la relación entre las dos fuentes fuese estable, la línea debería ser horizontal. Y lo que vemos es que el ingreso jubilatorio de EPH es cada vez más bajo que el que corresponde a los datos de ANSES.

Eso tiene muchas consecuencias. Por ejemplo, si medimos la pobreza en esos años, incluso usando una canasta actualizada con precios reales, nos puede dar un nivel de pobreza más alto de lo que era en realidad (siempre que el sesgo de los ingresos haya afectado también a los sectores de ingresos medios para abajo). Pero como contracara, también la desigualdad puede darnos más baja de lo que realmente era (en caso que los ingresos más altos estén menos representados).

El tema es que esto no sólo importa cuando nos ponemos a usar los datos de los años cuestionados. Porque, a medida que la EPH releve mejor los ingresos, estos van a perder ese sesgo que los mostraba más bajos de lo que eran. Entonces, a la variación que tengan los ingresos realmente, se les va a agregar un pequeño efecto distorsivo de aumento (así, por ejemplo, si los ingresos suben, va a parecer que la mejora es mayor de la que ocurrió en realidad; y si se deterioran, parte del deterioro no se va a reflejar en los datos). Eso es precisamente lo que vemos en los datos de la

EPH de 2016, que cuando comparamos con el último dato de 2015, nos muestran ingresos que crecieron más que la inflación, algo que definitivamente no fue lo que ocurrió (como lo demuestra la comparación entre los datos de salarios de asalariados registrados y de jubilaciones frente a la inflación). Y también podría tener que ver, al menos en parte, con ese abrupto aumento de la desigualdad que se observa desde 2016.

Continuidades y cambios entre 2015 y 2016

Otra forma de poner a prueba la idea de que los datos estaban ocultando un empeoramiento enorme de la realidad social, es compararlos con los de 2016. Si fuera cierta la suposición, deberíamos ver un empeoramiento abrupto. Sin embargo no es lo que ocurrió. De hecho en el caso de los ingresos (como acabamos de ver) fue al revés.

En el caso de la pobreza puede parecer que hubo un empeoramiento, pero eso es porque a veces no se compara la pobreza medida con la misma vara (o sea, con la misma canasta). Si calculamos la pobreza de 2015 con la canasta nueva (que empezó a usarse en 2016), vamos a ver que la pobreza del primer semestre de 2015 habría sido de entre 29% y 32%, (según qué referencia se utilice para considerar los precios); es decir, similar o incluso más alta que la primer medición semestral de 2016 (que fue de 30,3% en el segundo semestre). Y esto no se debe a que la pobreza se haya reducido, o mantenido idéntica, en ese periodo: muy por el contrario, los datos de fuentes administrativas muestran que se deterioró entre 4 y 8% la capacidad adquisitiva del salario promedio, de la jubilación mínima y de la Asignación Universal por Hijo (lo que haría esperar un incremento de la pobreza de entre 2 y 3 puntos porcentuales, tal como se verifica en otros periodos). La razón es que los datos de la EPH no tendían a ocultar la pobreza sino, en contra de lo que se suponía por aquel entonces, a exagerarla.

Una de las cuestiones más espinosas tiene que ver con la tasa de desocupación. Para esto no existen elementos para dar una respuesta certera. Si bien existen indicios de que el indicador pudo estar subestimado (lo que ocurriría por clasificar como inactivos –no trabajan ni buscan trabajo– a quienes en realidad eran desocupados

—no trabajan pero buscan trabajo activamente—), difícilmente la magnitud de esta alteración haya superado un punto porcentual (algo que en nuestro trabajo estimamos a partir de cómo evolucionó la cantidad de puestos de trabajo del sector formal según los registros administrativos, de los valores del INDEC desde 2016 en adelante, o de los datos de las estadísticas de CABA).

Las proyecciones

Como comentamos al principio, la EPH es una encuesta que no se realiza a toda la población sino a una muestra. Para el diseño de esa muestra resulta relevante la información sobre la estructura de la población, por edad, sexo y ubicación geográfica. Por dar un ejemplo exagerado, no va a ser igual una muestra para representar a una población total de 100 millones de personas de las cuales la mitad son personas mayores, que otra de 50 millones con una décima parte de personas mayores. Y esa información es importante saberla constantemente, porque cada año las poblaciones cambian a medida que personas crecen y algunas personas mueren y otras nacen.

Ahora supongamos que en algún momento, 2005 por ejemplo, detectamos que la muestra no es representativa. Y entonces la ajustamos. Por ejemplo, vemos que las personas mayores estaban subrepresentadas (estaban representadas como si fueran una menor parte de la que realmente eran) y hacemos el correspondiente ajuste. Desde 2006 usamos la muestra ajustada y el problema es que, cuando comparemos los datos de 2005 y 2006, una parte de las variaciones no va a tener que ver con que la realidad social se modificó, sino con que ajustamos la muestra. Así, por ejemplo, supongamos que de un año a otro la pobreza en una población se mantuvo idéntica pero entre las personas mayores la pobreza es más baja que en el promedio. Lo que va a pasar es que en ese caso hipotético en el que agregar personas mayores para ajustar la muestra, el dato de pobreza se va a modificar en la muestra nueva, y al comparar ambos años va a parecer que la pobreza se redujo.

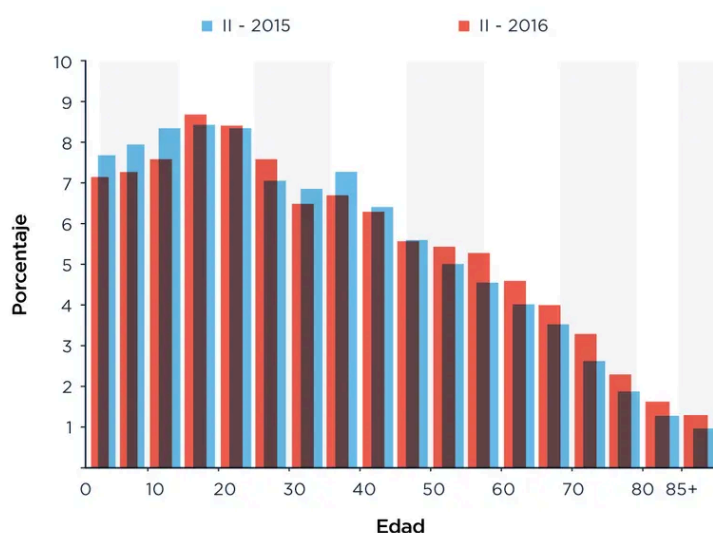
Con todo esto no estamos intentando imaginar contratiempos improbables, porque ajustar la muestra cada tanto es algo ineludible. El problema es que la información sobre la estructura de la población la tenemos, propiamente, sólo

cada 10 años, con cada censo de población. Entre un censo y otro lo que hacemos es realizar ‘proyecciones de población’. Y como difícilmente las proyecciones son exactas, el día que volvemos a tener información (del siguiente censo), casi siempre encontramos diferencias al comparar la población que creíamos tener con la que ahora sabemos que tenemos; se hace necesario ajustar la muestra. Entonces, luego de cada censo nacional de población, el INDEC corrige las proyecciones poblacionales que se habían elaborado a partir del censo anterior y construye una nueva muestra maestra para la EPH. Esto puede ser todo un dolor de cabeza si no avisan qué cambió, ni cuándo ni por qué: quienes usamos la información no vamos a saber cuánto de lo que vemos variar en los datos se explica por el ajuste de la muestra ni qué tan válido es comparar los datos de la muestra ajustada con los datos previos.

Sobre este tema, hay una crítica válida a la EPH para el periodo cuestionado: luego del Censo 2010 llegó el momento de ajustar el marco muestral, porque ahora teníamos datos que podían sustituir las proyecciones (hechas en base al Censo 2001). El problema es que se hizo sin transparencia. Entre 2013 y 2015, sin advertir que se haría ni explicar cómo, las proyecciones fueron sustituidas por los nuevos datos. Para los usuarios, conocer los cambios que eso supuso en la muestra, o su fundamento, implicó hacer inferencias, pero sin certezas.

Lo curioso es que, inesperadamente, encontramos que ocurrió algo muy similar a partir de 2016. Supuestamente (así lo informan oficialmente), la nueva gestión continúa utilizando las proyecciones y el marco muestral basados en el Censo 2010. Es decir, ese que con poca transparencia se introdujo entre 2013 y 2015. Sin embargo, nos encontramos con que desde 2016 hubo cambios en la estructura de población, con algunas diferencias llamativas respecto de la que veíamos en la muestra de 2015. Hubo modificaciones en la representación geográfica, de modo que los Partidos del Gran Buenos Aires pasaron a tener, desde 2016, un peso mucho mayor en el total (en detrimento de los aglomerados del Interior). También se modificó la estructura de edades: aumentó sustancialmente la participación de las personas mayores (en detrimento, sobre todo, de los menores de 15 años). No hubo advertencia ni justificación alguna para la introducción de

esos cambios. Y todo esto importa bastante porque la estructura de la población tiene impacto en indicadores como pobreza, desigualdad de ingresos y mercado de trabajo (por ejemplo, la tasa de pobreza es alrededor de cinco veces mayor entre los menores de 15 años que entre los mayores de 60, así que un aumento en la participación de las personas mayores va a reducir los niveles totales de pobreza).



Estructura de edades de la muestra expandida de la EPH II-2015 y II-2016

Claroscuros

Evidentemente, entre 2007 y 2015, la EPH tuvo problemas. Es grave que los haya tenido y graves fueron sus consecuencias sobre la confianza en los datos. Pero eso no alcanza para bajar los brazos. Echar luz sobre la realidad es una tarea compleja (por lo menos tan compleja como la realidad misma) y por eso **los problemas que actualmente pueda tener la EPH no implican que debemos desacreditarla sin más**. Lo cierto es que, aún con sus claroscuros, **la EPH ilumina mejor la realidad social de esos años de lo que puede hacer otra alternativa**. En el caso de la pobreza, por ejemplo, cualquier medición basada en la EPH –con precios de índices alternativos– resulta mucho más ‘realista’ que lo que informó la UCA (Universidad Católica Argentina) durante esos años.

O sea que la EPH no es ‘irrecuperable’. Lo que necesitamos es estudiar a fondo sus limitaciones y sesgos, producir más información, y no desistir. Reconstruir el

pasado es construir el futuro. Y matar al encuestador no produce mejores datos.

Nota de los autores: como científicos sociales que usamos la información estadística como uno de nuestros principales insumos, nos sentimos especialmente tocados con la desconfianza que afecta los datos oficiales que nos permiten analizar la realidad social. Por eso nos propusimos analizar en detalle la calidad de los datos y establecer cuáles se pueden usar y con qué precauciones. El resultado es un libro que acabamos de publicar, digital y de acceso abierto:



Claroscuros, 9 años de datos bajo sospechas. EPH en Argentina entre 2007 y 2015.

elgatoylacaja.com/no-maten-al-encuestador

Sumate en 
eglc.ar/bancar