



Consejo Nacional de Evaluación
de la Política de Desarrollo Social

Metodología para la imputación de ingresos de la ENOE en el marco de la estimación del porcentaje de pobreza laboral

Nota técnica

Diciembre de 2024

Contenido

Introducción	3
Los valores perdidos en las encuestas	5
Evolución de los datos perdidos en el ingreso declarado	7
Imputación de ingresos por rangos de salarios mínimos del CONEVAL	15
Propuesta de imputación de ingresos laborales	18
Exploración de variables	19
Imputación simple	22
Imputación múltiple.....	23
Resultados.....	26
Conclusiones	30
ANEXO 1	31
Métodos de imputación de ingresos	31
Experiencias Internacionales y en México	31
Experiencias internacionales	33
Propuestas en México	36
ANEXO 2	40
Variables exploradas	40
Bibliografía.....	43

Introducción

El Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) estima el porcentaje de la población en situación de pobreza laboral trimestralmente, a partir de la información publicada en la Encuesta Nacional de Ocupación y Empleo (ENOE) que genera el INEGI. Este indicador de corto plazo muestra la proporción de la población que, aún al hacer uso de todo su ingreso laboral disponible en su hogar no puede adquirir los productos de la canasta alimentaria, representada por la Línea de Pobreza Extrema por Ingresos (LPEI).¹

De esta forma, si el ingreso laboral aumenta más que el valor de la canasta alimentaria, el poder adquisitivo del hogar mejora y disminuye la pobreza laboral. La pobreza laboral es un indicador que sirve como herramienta preventiva de corto plazo del ingreso laboral de los hogares y su relación con la evolución del valor monetario de la canasta alimentaria, es decir, del poder adquisitivo del ingreso de los hogares.

Para dar seguimiento al poder adquisitivo del ingreso laboral en el tiempo, se requiere conocer cómo se estima el porcentaje de pobreza laboral y cuáles son los desafíos que se encuentran asociados a este proceso de medición. En este sentido, la metodología del Índice de la Tendencia Laboral de la Pobreza tiene por objetivo cuantificar la capacidad de los hogares para adquirir la canasta alimentaria con base en sus ingresos laborales. Sin embargo, el proceso de cálculo enfrenta ciertos retos debido a las limitaciones de las fuentes de información.

El cálculo del porcentaje de pobreza laboral consiste en obtener el ingreso laboral per cápita, es decir, sumar el ingreso laboral de los ocupados del hogar y dividirlo entre el total de integrantes, para posteriormente compararlo con la LPEI y conocer si es suficiente para obtener una canasta alimentaria para cada integrante. Sin embargo, la presencia de datos faltantes (*missing values*) del ingreso laboral en la ENOE representa un reto para el cálculo del indicador.² Por un lado, existe un porcentaje de la población que no reporta su ingreso laboral exacto, pero sí proporciona información sobre el rango en términos de número de salarios mínimos percibidos (a lo que se le denomina rango de salario mínimo), lo cual permite recuperar el valor del ingreso mediante un proceso denominado imputación. Por otra parte, también existe el caso de que el entrevistado no haya informado ni su ingreso exacto ni el rango de salario

¹ El porcentaje de pobreza laboral se emplea como insumo para generar el Índice de la Tendencia Laboral de la Pobreza (ITLP), este indicador muestra la evolución de la pobreza laboral respecto al periodo base (el primer trimestre de 2020), en el cual toma el valor de la unidad. El ITLP forma parte del Catálogo Nacional de Indicadores desde 2013.

² Los datos faltantes ocurren cuando el informante no responde a una o más preguntas y se deben principalmente a que el informante no sabe la respuesta o bien se niega a responder.

mínimo, por lo que no es posible calcular su ingreso laboral per cápita y, por ende, tampoco es posible identificar su situación de pobreza laboral.

Dada esta limitante y con el objetivo de contar con un indicador para dar seguimiento en el tiempo, el CONEVAL optó por excluir de los cálculos las observaciones para las que no es posible estimar la situación de pobreza laboral. En este sentido, se identifican a los ocupados que no tienen valores válidos de ingreso (es decir, no reportan su ingreso puntual ni el rango de salario mínimo), y se elimina a los hogares que tienen al menos un ocupado en esta situación.

Estos hogares no se contabilizan en la estimación del total de la población con ingreso laboral per cápita inferior a la canasta alimentaria ni en el total de la población. En consecuencia, la población de estos hogares se excluye tanto del numerador como del denominador del indicador del porcentaje de pobreza laboral. Así, el CONEVAL presenta las estimaciones de pobreza laboral en términos relativos (porcentaje de la población en pobreza laboral), ya que mostrar el número absoluto de personas en pobreza laboral se ve limitado por la falta de información completa para toda la población.

Diversos estudios han mencionado que el método de recuperación de ingresos mediante rangos de salario mínimo, implementado por el CONEVAL, cuenta con áreas de oportunidad, ya que existen herramientas para el tratamiento a los valores perdidos que podrían incluirse en las estimaciones. Un adecuado tratamiento de valores perdidos es requerido debido a que se ha encontrado que la falta de reporte de ingresos laborales en la ENOE no sigue una distribución aleatoria, sino que está relacionada con algunas características de los ocupados, como el nivel de instrucción (Campos, 2013; CONASAMI, 2023); es decir, se observa una correlación entre la proporción de valores perdidos y un mayor nivel de instrucción.

En este sentido, un apropiado método para imputar los valores faltantes de ingreso laboral tiene tres beneficios: primero, permite disminuir el sesgo de subestimación del ingreso laboral; segundo, mejora el cálculo de la variabilidad de las estimaciones; y, tercero, la imputación de los ingresos faltantes posibilita identificar la situación de pobreza laboral para la población sin información de ingreso laboral, de tal forma que permite estimar la pobreza laboral tanto en términos relativos como absolutos (total del número de personas); estos puntos se abordan posteriormente en este documento.

El objetivo del presente documento es proponer un método de imputación de ingresos adecuado para la estimación de pobreza laboral que minimice los sesgos en el cálculo

del ingreso como de su variabilidad y que se adecue alinea con los principios de transparencia, objetividad y rigor técnico, por lo que se presenta la revisión de distintas metodologías planteadas a nivel nacional e internacional y la propuesta de ejercicios de imputación de ingresos con sus respectivas estimaciones de pobreza laboral.

Con esta finalidad, la estructura del presente documento se compone por seis apartados: primero se expone la presencia de valores perdidos en las encuestas de hogares y los tipos en los cuales pueden ser clasificados; en segundo lugar, se explora la evolución y caracterización de los datos perdidos en el ingreso laboral de la ENOE; posteriormente, se aborda la metodología empleada hasta ahora por el CONEVAL para la imputación de ingreso por rango de salarios mínimos en la estimación de pobreza laboral; enseguida, se presentan las metodologías de las propuestas de métodos de imputación; posteriormente, se presentan los resultados para pobreza e ingreso laboral aplicando dichas metodologías; finalizando con reflexiones finales acerca de los resultados obtenidos y puntualizando los siguientes pasos.

Los valores perdidos en las encuestas

La ENOE,³ elaborada por el INEGI, es el proyecto estadístico continuo más grande de México (INEGI, 2024), este proporciona información mensual y trimestral sobre la situación laboral de las y los trabajadores, lo cual permite conocer las características del mercado laboral.

El INEGI publica trimestralmente los microdatos de la ENOE, a partir de los que es posible estimar indicadores asociados al mercado laboral a nivel nacional, entidad, ámbito de residencia y ciudad autorrepresentada, a su vez, el CONEVAL retoma los microdatos de la ENOE para estimar el indicador de pobreza laboral. En este apartado se ahondará en algunas limitaciones inherentes en el uso de la información de las encuestas como la ENOE, específicamente la ausencia de respuesta a preguntas que da lugar a valores perdidos en las bases de datos.

La ausencia de respuesta a preguntas sobre ingresos en encuestas es un fenómeno común que afecta la calidad y la fiabilidad de los datos recolectados debido a que la mayoría de los procedimientos estadísticos no están diseñados para manejar esta

³ Se hace referencia tanto a la ENOE como a la ENOE^N. En septiembre de 2020 el INEGI dio a conocer la ENOE^N que incorpora dos estrategias de levantamiento: entrevistas cara a cara y vía telefónica. La publicación de la ENOE^N ocurrió después de haberse interrumpido la ENOE del segundo trimestre 2020 debido a la contingencia sanitaria por la COVID-19. Asimismo, en el primer trimestre de 2023, se retomó de manera regular la ENOE, bajo una estrategia operativa mixta de entrevistas cara a cara y telefónicas. Para más información se sugiere consultar la documentación de la ENOE del INEGI, disponible en: <https://www.inegi.org.mx/programas/enoe/15ymas/>.

situación. Este fenómeno se presenta cuando los encuestados se niegan o no pueden proporcionar información sobre sus ingresos, ya sea por razones de confidencialidad, desconocimiento o percepción de sensibilidad de la información solicitada (Schafer & Graham, 2002; Schräpler, 2004).

Los datos faltantes tienden a ubicarse en los extremos de la distribución de ingresos, situación que complica aún más el análisis de los ingresos. Además, factores como el diseño del cuestionario, la formulación de las preguntas y las características sociodemográficas del encuestado y del entrevistador influyen significativamente en la propensión a no responder (Frick & Grabka, 2005; Schräpler, 2004).

Adicionalmente, existen conceptos clave como los mecanismos de datos faltantes: completamente al azar (MCAR), al azar (MAR) y no al azar (MNAR). Estos mecanismos definen la relación entre la probabilidad de que un dato esté faltante y los valores de los datos observados y no observados. MCAR implica que la falta de datos no depende de los valores de los datos, mientras que MAR permite la dependencia de los datos observados, pero no de los no observados. MNAR sugiere que la probabilidad de que un dato esté faltante depende de los propios valores faltantes, lo cual complica significativamente el análisis (Schafer & Graham, 2002; Si & Reiter, 2013). De acuerdo con lo anterior, se observa que los datos faltantes del ingreso laboral de la ENOE tienen un comportamiento MAR, es decir, sus valores no son del todo aleatorios, sino una parte depende de una serie de características de las personas informantes, en este caso, de las personas ocupadas.

La falta de respuesta a algunas preguntas en encuestas es común, además se reconoce que este fenómeno no es aleatorio dentro de estos instrumentos, en específico cuando se trata de la variable de ingresos en la ENOE. Por lo que, no sólo existe la metodología de imputación por rangos de salarios mínimos que actualmente lleva a cabo el CONEVAL, sino tanto en México como en diversos países se han implementado otras metodologías para imputar los valores perdidos con el objetivo de ofrecer a los usuarios información completa y así disminuir el sesgo producido en el análisis de la información por eliminar los casos faltantes; por lo que, en el Anexo 1 se encuentra la revisión de los métodos de imputación de ingresos y las experiencias internacionales y en México. En el siguiente apartado se presenta la exploración de los datos faltantes del ingreso laboral en la ENOE.

Evolución de los datos perdidos en el ingreso declarado

El reporte del ingreso laboral capturado con la ENOE puede clasificarse en uno de los siguientes tres tipos: i) reporte de valores positivos, ii) reporte de valores cero (los cuales corresponden a las personas ocupadas no remuneradas) o iii) ingreso no reportado o declarado (*missing values*). Para este último tipo, que en adelante se hará referencia como *missing values (mv)*, existen casos en las que es posible recuperar el valor del ingreso no declarado a través de la clasificación que hace el informante en uno de los siete rangos de número de salarios mínimos, en adelante se referirá a estos casos como *missing values con rango de salario mínimo (mvcrsm)*, en el Cuadro 1. se presentan los rangos de salario mínimo y la distribución de las observaciones; sin embargo, existen otros casos en los que no es posible recuperar el ingreso laboral, ya que el informante, que se declaró como ocupado con remuneración, no reportó su ingreso ni lo clasificó en alguno de los rangos de salario mínimo, en adelante se referirá a estos casos como *missing values sin rango de salario mínimo (mvsrsm)*.

Cuadro 1.

Número y porcentaje de observaciones de personas ocupadas que reportaron ingreso laboral a través de algún rango de salario mínimo (*mvcrsm*)

ENOE – muestral

Tercer trimestre 2024

Rango de salario mínimo	Observaciones	%
Menor que uno	5,824	21.9
Igual que uno	7,412	27.9
Más de 1 hasta 2	9,380	35.2
Más de 2 hasta 3	2,914	11.0
Más de 3 hasta 5	713	2.7
Más de 5 hasta 10	293	1.1
Más de 10	79	0.3
Total	26,615	100

Fuente: estimaciones del CONEVAL con base en la ENOE del INEGI.

Nota: se excluyen casos no completos que son aquellas observaciones ocupadas que no especifican su categoría en las variables de: Nivel de instrucción, Duración de la jornada y Condición de ocupación.

Derivado de ello, resulta indispensable identificar si estos valores perdidos de ingreso se dan de manera aleatoria en las personas ocupadas o si existen ciertas características de estas que hagan más probable que no reporte su ingreso en las encuestas. Conocer la distribución de los valores perdidos de ingreso entre los ocupados permite elegir el método de imputación óptimo para mejorar la medición de la pobreza laboral minimizando los sesgos.

En este sentido y considerando los dos tipos de *missing values*, se exploró el mecanismo de comportamiento de datos faltantes, para lo que se realizó un análisis sobre las características socioeconómicas de las observaciones con *missing values*. Como lo muestra la Gráfica 1, del primer trimestre de 2005, periodo de inicio de la serie de la ENOE, al tercer trimestre de 2024, periodo más reciente con el que se cuenta información al momento de realizar esta nota, es posible identificar un aumento de las observaciones que no reportaron directamente su ingreso laboral, es decir, de los *mvcrsm*, así como de los *mvsrsm*. Al respecto, se observa que, a nivel muestral, al inicio de la serie de la ENOE, es decir, en el primer trimestre de 2005, el 4.6% del total de ocupados eran *mvsrsm*, esta proporción aumentó 9.1 puntos porcentuales para el tercer trimestre de 2024, donde se presentó un valor de 13.7%.

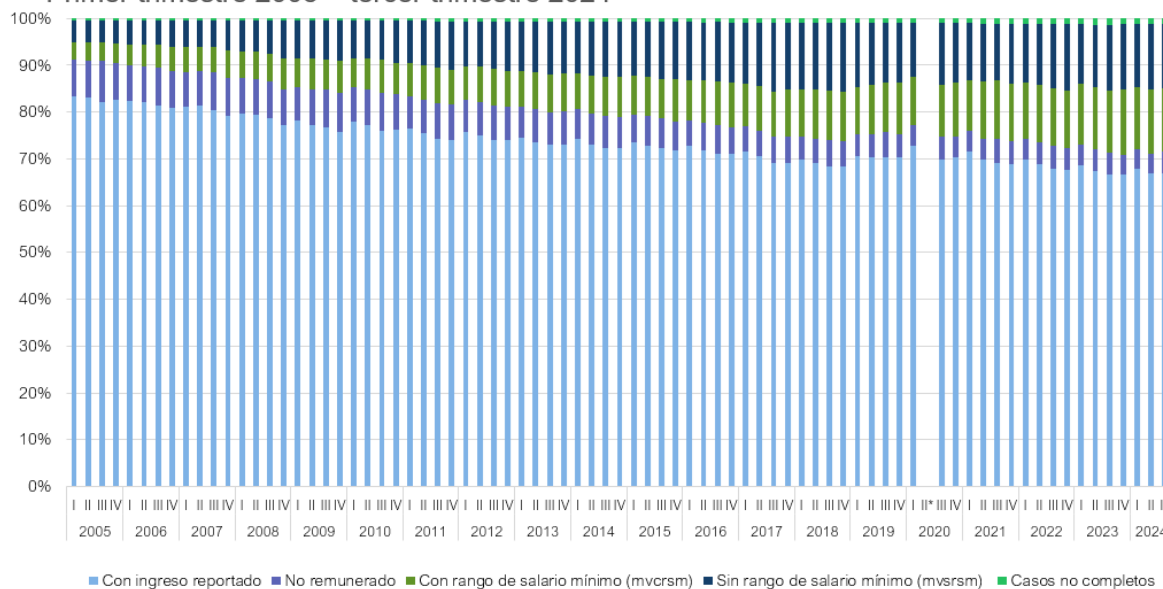
Por su parte, los *mvcrsm* presentaron un comportamiento similar, ya que entre el primer trimestre de 2005 y el tercer trimestre de 2024, la proporción de estas observaciones respecto al total de ocupados aumentó 10 puntos porcentuales, al pasar de 3.6% a 13.6%; es decir, en el tercer trimestre de 2024, este conjunto fue 3.8 veces el reportado en el primer trimestre de 2005.

Gráfica 1.

Porcentaje de observaciones de personas ocupadas según reporte de ingreso laboral

ENOE – muestral

Primer trimestre 2005 – tercer trimestre 2024



Fuente: estimaciones del CONEVAL con base en la ENOE y ENOE Nueva Edición (ENOE[®]). *Debido a la contingencia sanitaria por la COVID-19, el INEGI suspendió la recolección de información de la ENOE referente al segundo trimestre de 2020, por lo cual no se cuenta con el insumo necesario para el cálculo de los indicadores correspondientes a este periodo.

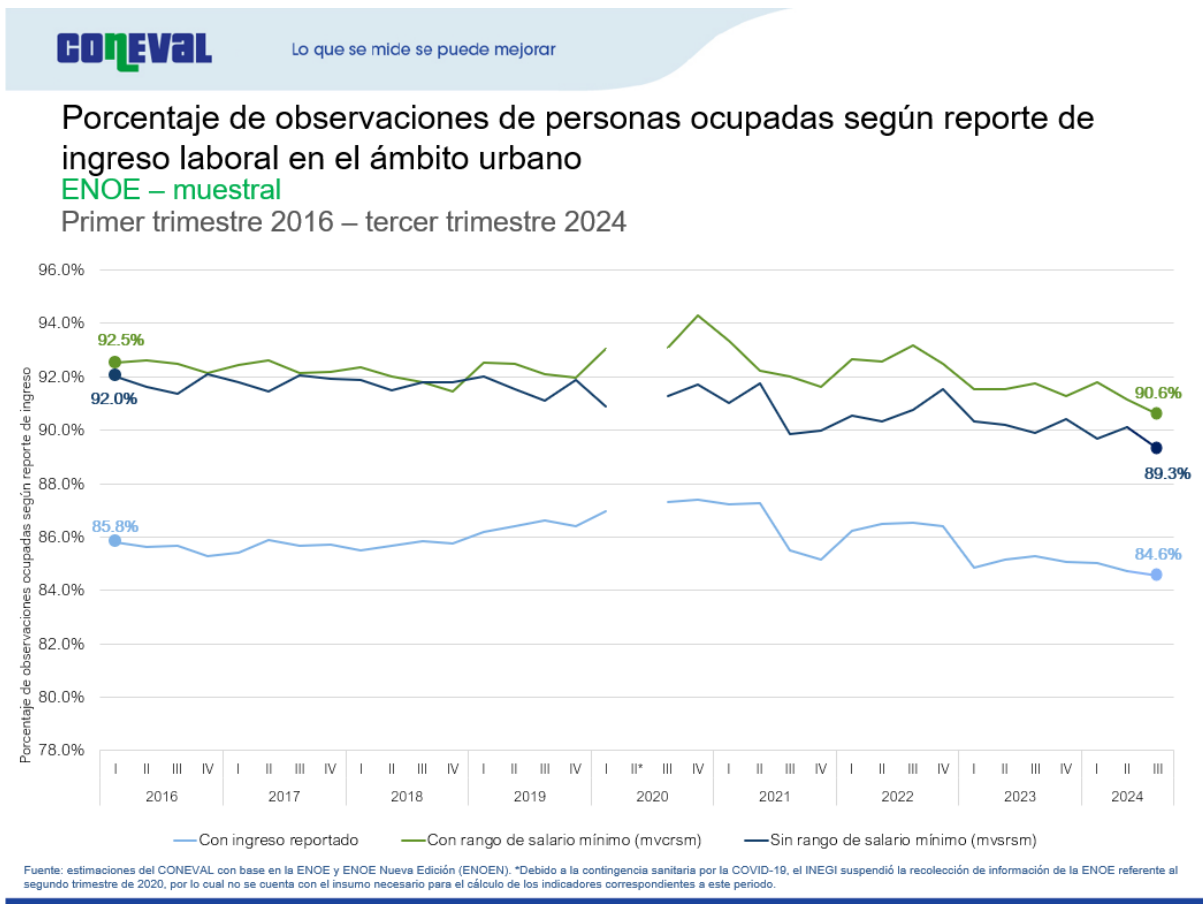
Nota: casos no completos se refiere a las observaciones ocupadas que no especifican su categoría en las variables de: Nivel de instrucción, Duración de la jornada y Condición de ocupación.

Adicional a dicho aumento, también es posible identificar algunas características de las observaciones con ingreso no reportado (*missing values*),⁴ lo que nos permitiría concluir que la falta de reporte del ingreso laboral no es aleatoria. Dentro de estas características, resalta que del total de *mvsrsm*, para el tercer trimestre de 2024, el 60.6% eran hombres, en este mismo sentido, del total de *mvcrsm* el 59.7% eran hombres; sin embargo, esta prevalencia es congruente con la tasa de ocupación nacional, la cual es mayor para los hombres en comparación con las mujeres.

Para el mismo periodo, del total de *mvsrsm*, se encontró que el 89.3% pertenecían al ámbito urbano, en comparación con el 10.7% de las observaciones del ámbito rural, tendencias que se han mantenido a lo largo de la serie. Adicionalmente, destaca que esta proporción es ligeramente mayor en los *mvcrsm*. Lo anterior se puede observar en la Gráfica 2.

⁴ Las observaciones con ingreso no reportado consideran a: observaciones de ingreso con rango de salario mínimo y observaciones de ingreso sin rango de salario mínimo.

Gráfica 2.



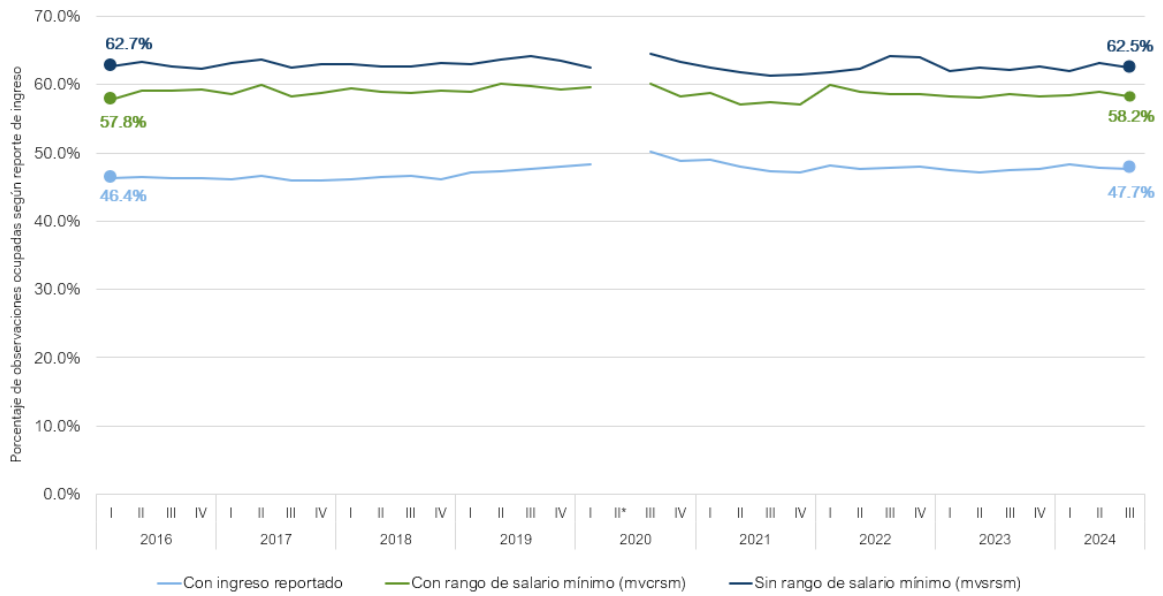
Otra de las características cuya proporción es mayor en los *mvrsrm* es la condición de formalidad, lo que se traduce en que la mayoría de las personas que no declaran su ingreso laboral están empleadas en el sector formal, contrario a la tasa reportada para todos los ocupados. En el tercer trimestre de 2024, del total de observaciones ocupadas con ingresos reportados el 47.7% se encontraban en algún empleo formal, mientras que la incidencia para los *mvcrsm* y los *mvrsrm* fue de 58.2% y 62.5%, respectivamente.

Gráfica 3.

Porcentaje de observaciones de personas ocupadas según reporte de ingreso laboral en condición de formalidad

ENOE – muestral

Primer trimestre 2016 – tercer trimestre 2024



Fuente: estimaciones del CONEVAL con base en la ENOE y ENOE Nueva Edición (ENOEN). *Debido a la contingencia sanitaria por la COVID-19, el INEGI suspendió la recolección de información de la ENOE referente al segundo trimestre de 2020, por lo cual no se cuenta con el insumo necesario para el cálculo de los indicadores correspondientes a este periodo.

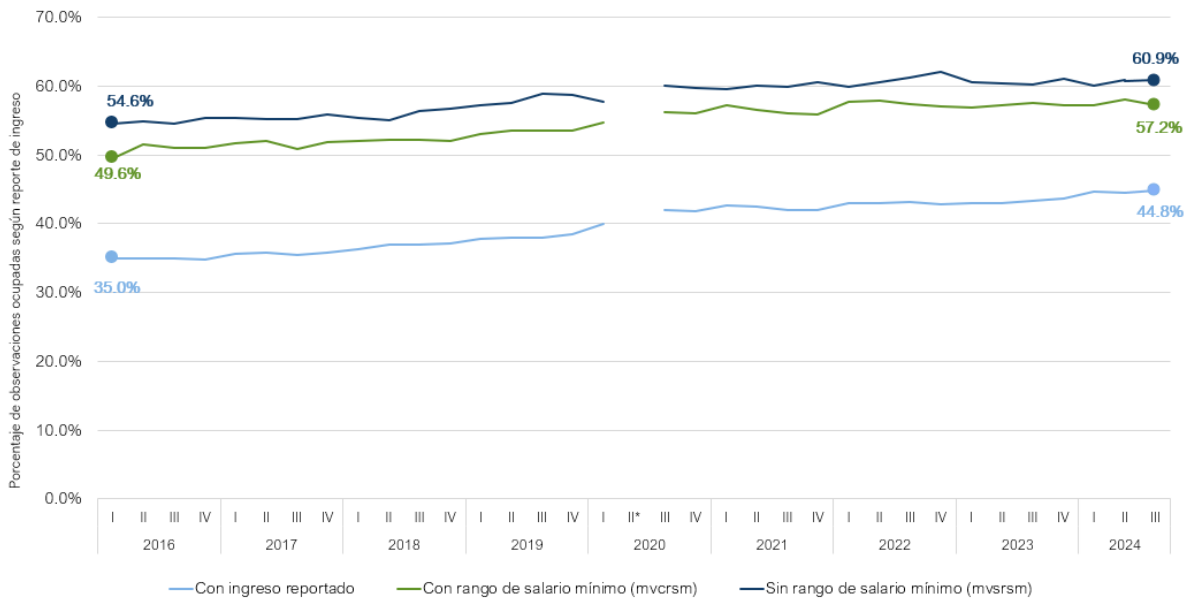
Asimismo, para el tercer trimestre de 2024, el 44.8% de las observaciones ocupadas (con remuneración) y que reportaron ingreso contaban con nivel de instrucción media superior o superior, en comparación con la proporción de los *mvcrsm* y los *mvrsrm*, las cuales rebasan el 50% del total de las observaciones desde 2016.

Gráfica 4.

Porcentaje de observaciones de personas ocupadas según reporte de ingreso laboral con nivel de instrucción media superior o superior

ENOE – muestral

Primer trimestre 2016 – tercer trimestre 2024



Fuente: estimaciones del CONEVAL con base en la ENOE y ENOE Nueva Edición (ENOEN). *Debido a la contingencia sanitaria por la COVID-19, el INEGI suspendió la recolección de información de la ENOE referente al segundo trimestre de 2020, por lo cual no se cuenta con el insumo necesario para el cálculo de los indicadores correspondientes a este periodo.

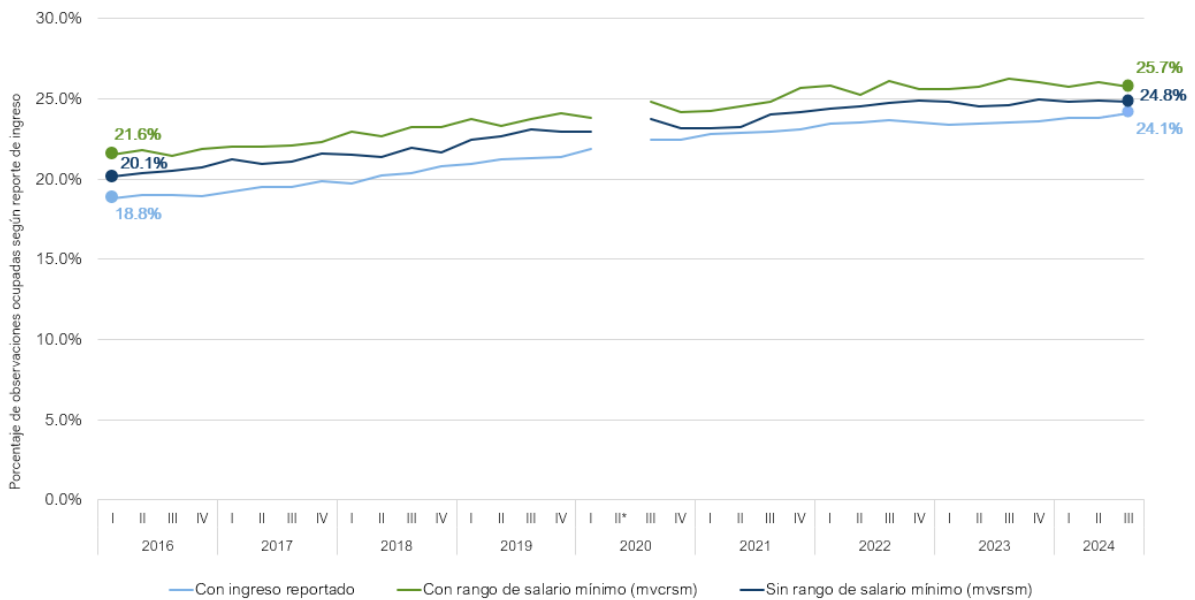
Si se analiza el nivel de instrucción separado por educación media superior y superior, se observa que, a lo largo de la serie en ambos niveles, la proporción de observaciones de los *mvcrsm* y los *mvrsrm* es mayor que la proporción que presentan las observaciones con ingreso reportado. Específicamente, respecto al nivel de instrucción media superior, se observa que la proporción mayor es la de los *mvcrsm*; mientras que, en el nivel de instrucción superior la mayor proporción corresponde a los *mvrsrm*; sin embargo, es importante destacar que se presenta un mayor porcentaje en las observaciones tanto de los *mvcrsm* y los *mvrsrm* que tienen un nivel de instrucción superior en comparación con aquellas que tienen un nivel medio superior.

Gráfica 5.

Porcentaje de observaciones de personas ocupadas según reporte de ingreso laboral con nivel de instrucción media superior

ENOE – muestral

Primer trimestre 2016 – tercer trimestre 2024



Fuente: estimaciones del CONEVAL con base en la ENOE y ENOE Nueva Edición (ENOEN). *Debido a la contingencia sanitaria por la COVID-19, el INEGI suspendió la recolección de información de la ENOE referente al segundo trimestre de 2020, por lo cual no se cuenta con el insumo necesario para el cálculo de los indicadores correspondientes a este periodo.

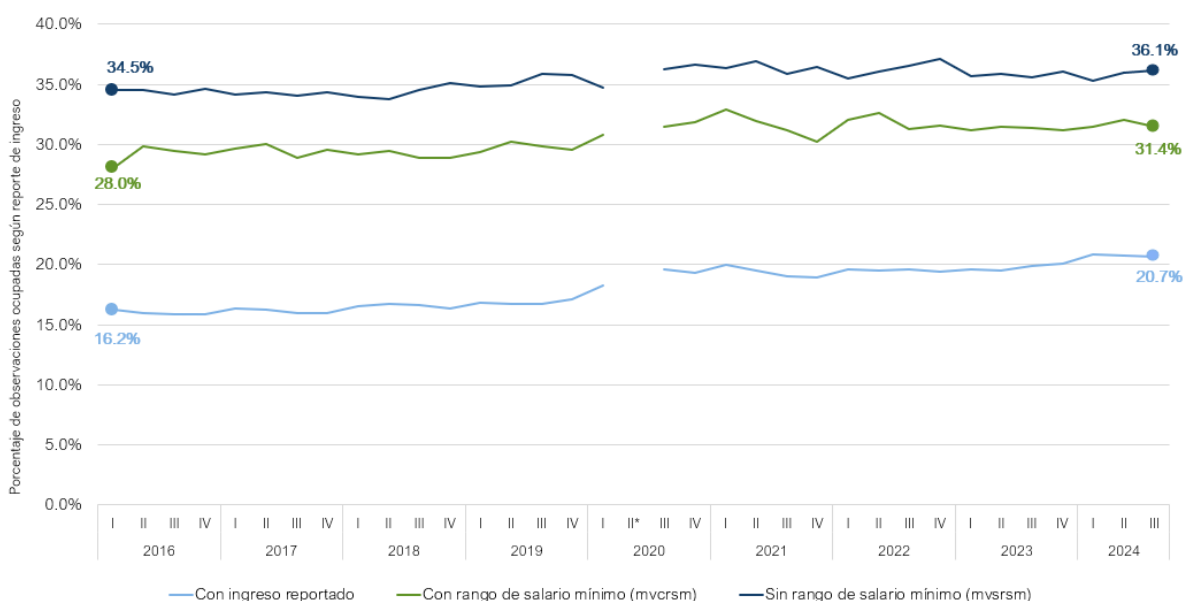
Para este último nivel de instrucción (educación superior) resalta que, para el tercer trimestre de 2024, la proporción de los *mvsrsm* fue 15.5 puntos porcentuales mayor a la presentada en las observaciones con ingreso reportado, esta tendencia se mantuvo a lo largo de la serie, tal y como lo muestra la Gráfica 6.

Gráfica 6.

Porcentaje de observaciones de personas ocupadas según reporte de ingreso laboral con nivel de instrucción superior

ENOE – muestral

Primer trimestre 2016 – tercer trimestre 2024



Fuente: estimaciones del CONSEVAL con base en la ENOE y ENOE Nueva Edición (ENOEN). *Debido a la contingencia sanitaria por la COVID-19, el INEGI suspendió la recolección de información de la ENOE referente al segundo trimestre de 2020, por lo cual no se cuenta con el insumo necesario para el cálculo de los indicadores correspondientes a este periodo.

Lo anterior refleja que la mayoría de las observaciones que no declaran su ingreso laboral y el cual no es posible recuperar a través de algún rango de salario mínimo, son hombres que habitan en el ámbito urbano, con nivel de instrucción media superior o superior y que están ocupadas en actividades formales, lo que nos permite intuir que estos valores perdidos no se dan de forma aleatoria. En este mismo sentido, con el fin de reforzar la hipótesis de que ambos conjuntos de observaciones (las que declaran ingreso y las que no) cuentan con características diferentes, se realizó una serie de diferencias de medias con lo cual se confirmó que estas características muestran diferencias estadísticamente significativas, cuyos resultados pueden ser consultados en el Cuadro 2.

Cuadro 2.

Prueba de diferencia de medias en población ocupada según reporte de ingreso laboral ENOE – muestral Tercer trimestre 2024

Características	Reporta	No reporta ^{1/}	Diferencia	Estadístico t
Hombre	0.5802	0.6015	-0.0213	-8.4021 ***
Urbano	0.8456	0.8994	-0.0538	-30.3443 ***
Empleo formal	0.4773	0.6039	-0.1266	-49.6163 ***
Medio superior y superior	0.4479	0.5906	-0.1427	-56.0045 ***
Duración de la jornada de 35 a 48 horas	0.4902	0.5394	-0.0492	-18.8484 ***
Empresas, Gobierno e Instituciones	0.5965	0.7146	-0.1182	-47.9405 ***
Trabajadores subordinados y remunerados	0.7551	0.7095	0.0456	20.2885 ***
Profesionales, técnicos y trabajadores del arte	0.1060	0.1786	-0.0725	-42.5962 ***
Empresas y negocios	0.5420	0.5918	-0.0499	-19.5305 ***
Total de observaciones	130,527	53,343		

Ho: existe diferencia en las medias entre los dos grupos de observaciones declaradas ocupada.

Fuente: estimaciones del CONEVAL con base en la ENOE del INEGI. ^{1/} Se refiere a la muestra de población ocupada que no reportó ingreso laboral o que lo reportó a través de algún rango de salario mínimo. Nota: la muestra de población ocupada incluye observaciones de 12 años y más. Se excluyen casos no completos que son aquellas observaciones ocupadas que no especifican su categoría en las variables de: Nivel de instrucción, Duración de la jornada y Condición de ocupación. (***) estadísticamente significativo a un nivel de confianza del 95%.

Imputación de ingresos por rangos de salarios mínimos del CONEVAL⁵

Existen varios métodos para recuperar la información faltante en materia de ingreso obtenida con la ENOE, en el caso de CONEVAL, el rescate de la información se hace a través de la imputación por rangos de salarios mínimos, la cual, como se mencionó, forma parte de la estimación actual de la pobreza laboral que actualiza trimestralmente.

El porcentaje de la población en situación de pobreza laboral hace referencia a la población con ingreso laboral inferior al valor monetario de la canasta alimentaria (Línea de Pobreza Extrema por Ingresos). El proceso para estimar este indicador

⁵ El CONEVAL estima trimestralmente el Índice de la Tendencia Laboral de la Pobreza (ITLP), el cual retoma como fuente de información a la ENOE del INEGI. Este indicador, que desde 2013 forma parte del Catálogo Nacional de Indicadores, permite dar seguimiento a la evolución del ingreso laboral y su relación con el valor monetario de la canasta alimentaria, es decir, obtener el porcentaje de la población en situación de pobreza.

emplea como insumo las bases de datos del Cuestionario sociodemográfico (SDEM) y del Cuestionario de ocupación y empleo ampliado parte II (COE2) de la ENOE.

En un primer paso, se recuperan los ingresos reportados de forma directa por los informantes (variable p6b2) de la base COE2, si es trabajadora o trabajador sin pago se asigna un ingreso cero; sin embargo, si no reportó el ingreso directo en la variable p6b2 y no es trabajadora o trabajador sin pago, se busca recuperar el ingreso que percibe a través del rango de salario mínimo declarado (variable p6c).

De tal forma que, el CONEVAL asigna el valor medio del rango de salario que declara el informante en la variable p6c, considerando el salario mínimo de cada una de las zonas de salario mínimo existentes,⁶ el cual se encuentra contenido en la variable salario de la base SDEM. De manera que, se asigna:

- Medio salario mínimo si reportó que el ingreso que percibe es menor que el valor vigente del salario mínimo (p6c = 1).
- Un salario mínimo si reportó que el ingreso que recibe es igual que el valor vigente del salario mínimo (p6c = 2).
- Un salario mínimo y medio si respondió que el ingreso que recibe es mayor que un salario mínimo y hasta dos (p6c = 3).
- Dos salarios mínimos y medio si reportó que el ingreso que percibe es equivalente a más de dos salarios mínimos y hasta tres (p6c = 4).
- Cuatro salarios mínimos si respondió que el ingreso que percibe es equivalente a más de tres salarios mínimos y hasta cinco (p6c = 5).
- Siete salarios mínimos y medio si reportó que el ingreso que percibe es equivalente a más de cinco salarios mínimos y hasta diez (p6c = 6).
- Diez salarios mínimos si reportó que el ingreso que percibe es mayor a lo que equivalen más de diez salarios mínimos (p6c = 6).

Finalmente, si el informante no reportó ingresos directamente en la variable p6b2, no indicó cuál es su ingreso mediante rangos de salario mínimo en variable p6c y no es trabajador o trabajadora sin pago, se asume como dato perdido que no es posible recuperar. La falta de información del ingreso laboral de al menos una persona en el hogar imposibilita calcular el total de su ingreso laboral, por lo que, no se consideran los hogares con al menos un integrante con ingreso laboral faltante en las estimaciones del porcentaje de la población en situación de pobreza. El proceso anterior se muestra en el siguiente esquema.

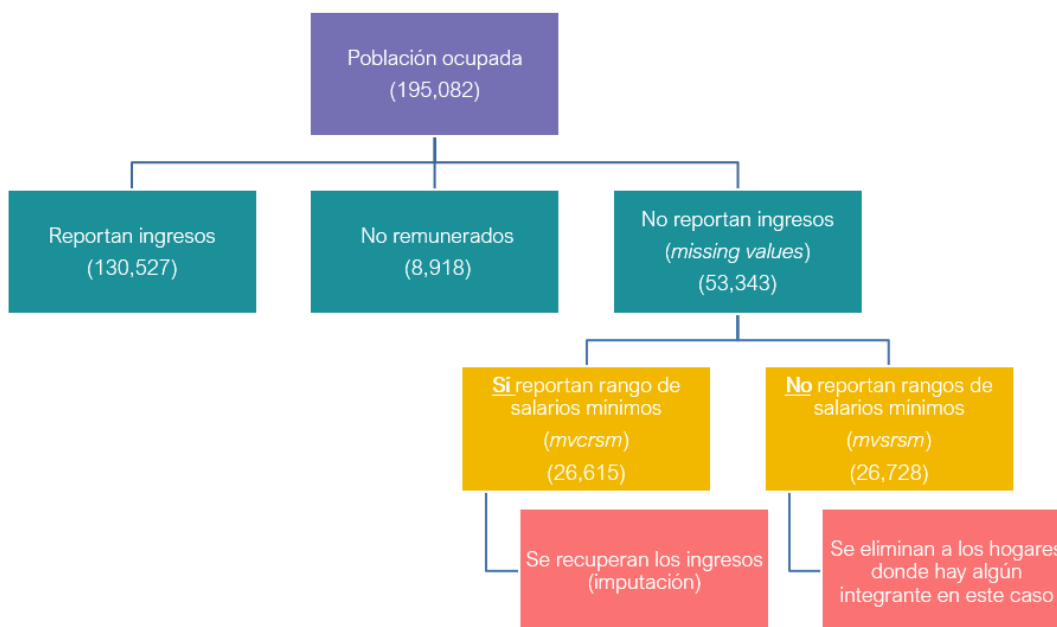
⁶ El consejo de representantes de la CONASAMI determina la distribución por áreas geográficas de los Salarios Mínimos Generales y Profesionales vigentes en el territorio nacional. Para más información, se sugiere consultar: <https://www.gob.mx/conasami/documentos/clasificacion-de-los-municipios-por-area-geografica>

Esquema 1.

Clasificación del ingreso reportado en la ENOE

Población ocupada

Tercer trimestre de 2024



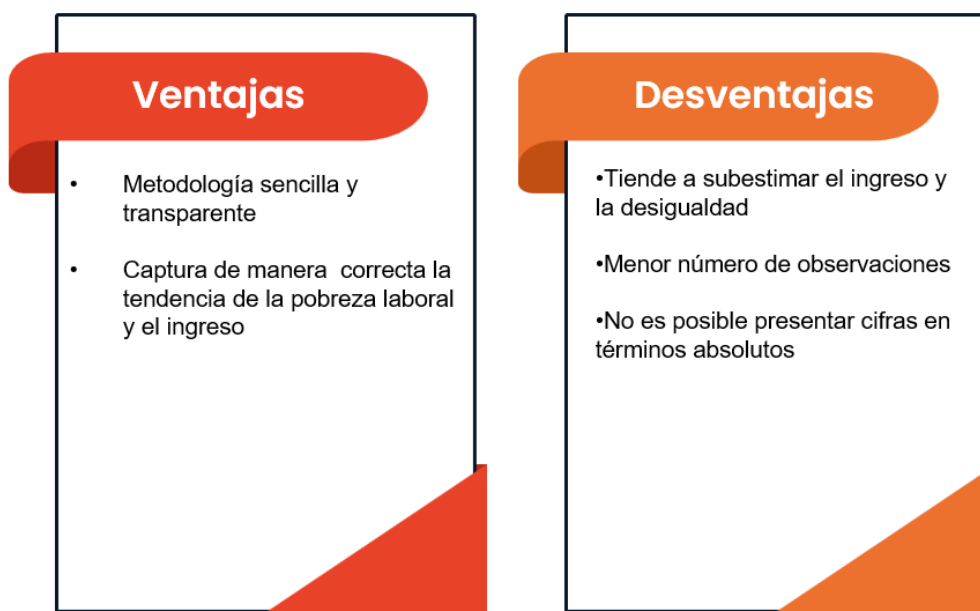
Fuente: estimaciones del CONEVAL con base en la ENOE.

Nota: se excluyen casos no completos que son aquellas observaciones ocupadas que no especifican su categoría en las variables de: Nivel de instrucción, Duración de la jornada y Condición de ocupación.

Sin embargo, como cualquier otro método, este posee algunas desventajas, la más importante de ellas es que se puede subestimar el ingreso y a su vez, sobreestimar el porcentaje de la población en pobreza laboral; asimismo, esta metodología no permite presentar cifras en términos absolutos. Lo anterior indica la necesidad de otras metodologías que permitan imputar los valores faltantes del ingreso laboral para que se cuente con una medición de pobreza laboral con un menor sesgo y recuperando en mayor medida la variabilidad de la distribución del ingreso laboral, la cual, a su vez, funja como insumo para el diseño de políticas públicas enfocadas a combatir la precariedad de salarios.

Cuadro 3.

Recuperación de ingreso laboral por intervalo de salarios



Fuente: elaboración del CONEVAL

Propuesta de imputación de ingresos laborales

El CONEVAL realizó una exploración de diferentes metodologías de imputación de ingresos en encuestas de hogares realizadas por organismos nacionales como internacionales, así como por la academia, esta revisión se presenta en el Anexo 1. Posteriormente, se seleccionaron aquellos métodos cuales podrían resultar adecuados para la imputación de ingreso laborales en la ENOE, con el objetivo final de ser utilizados para las estimaciones de pobreza laboral. Se buscó seleccionar aquellos métodos que permitieran estimar los ingresos faltantes dada la gama de información socioeconómica de la ENOE y reducir el sesgo de las estimaciones de pobreza laboral; adicionalmente, que el proceso fuera técnicamente robusto y transparente para su replicabilidad.

Para la elaboración de estas propuestas se consideran los tipos de valores faltantes (*mvcrsm* y *mvsrsm*) y se retomaron procesos de imputación simple y múltiple, de tal forma, que se pudiera contar con diferentes aproximaciones de imputación para contrastar los resultados. En cuanto a los modelos de imputación simple seleccionados, se probaron dos variantes: por un lado, imputación a partir de las medias del ingreso laboral de grupos con características similares; por otro lado, a partir de las medianas del ingreso laboral de grupos con características similares. En cuanto a los modelos de imputación múltiple se realizaron ejercicios con tres variantes: el método de imputación múltiple de ecuaciones encadenadas con paramiento de medias predictivas y con bosques aleatorios, y el algoritmo de imputación múltiple de HMISC.

En este apartado, se describe, en primer lugar, la exploración de variables socioeconómicas a utilizar como covariables para la imputación del ingreso; en segundo lugar, se exponen los métodos de imputación simple (medias y medianas) propuestos; en tercer lugar, se presentan los modelos de imputación múltiple utilizados.

Exploración de variables

El primer paso para la implementación de los métodos para la imputación de los ingresos se realizó una exploración de las variables a utilizarse en los modelos. Para ello se consideró una selección de 13 variables socioeconómicas de la ENOE, para mayor detalle ver la sección de variables exploradas del Anexo 2.

Esquema 2.

Preselección de variables ENOE-ENOE^N



Fuente: elaboración del CONEVAL.

Para reducir el número de variables que se preseleccionaron derivado de la revisión de la bibliografía y el análisis exploratorio previo, se realizaron modelos de árboles de regresión y clasificación (*Classification and Regression Trees, CART*), donde, para el caso del árbol de regresión la variable dependiente fue el ingreso y, para el árbol de clasificación fue la pobreza laboral.

Estos modelos se basan en algoritmos de aprendizaje, que abonan a la predicción o explicación de forma sencilla y representado mediante un árbol con las diversas combinaciones de las variables explicativas (nodos), mismas que pueden ser numéricas o categóricas. Para ello se definió una muestra de entrenamiento, para ambos modelos se seleccionó el 20% de la muestra para el entrenamiento y la muestra completa para la validación.

Estos modelos se realizaron para toda la serie de la ENOE, además con el objetivo de identificar con mayor facilidad cuáles eran las variables que explican mayormente el ingreso y la pobreza laboral, se retomó el *importance value* y se estandarizó a los

valores de importancia del trimestre, así los valores de importancia oscilan entre 0 y 100. Por lo que, finalmente se seleccionaron las variables con mayor importancia relativa a lo largo de la serie tanto para el ingreso como para pobreza laboral.

De esta forma se definió dos conjuntos de variables para los métodos. En el caso de los modelos de imputación simple de medias y medianas del ingreso se consideran las variables: clasificación de la población ocupada por posición en la ocupación (*pos_ocu*), clasificación de empleos formales e informales de la primera actividad (*emp_ppal*), filas agregadas de la matriz de Husmanns (*mh_fil2*), clasificación de la población cuenta propia en actividades no calificadas (*cp_anoc*), clasificación de la población ocupada por duración de la jornada (*dur_est2*)⁷ y clasificación de la población de 5 años y más por nivel de instrucción (*niv_ins2*)⁸ y zona salarial (*zona*).

Para el caso de los modelos de imputación múltiple se contemplan las variables: sexo de la persona (*sex*), clasificación de la población de 15 años y más por grupo de edad siete claves (*eda7c*), clasificación de la población de 5 años y más por nivel de instrucción (*niv_ins3*)⁹, clasificación de la población ocupada por posición en la ocupación (*pos_ocu*), clasificación de la población ocupada por condición de ocupación (*c_ocu11c*), clasificación de la población ocupada por duración de la jornada (*dur_est*), clasificación de empleos formales e informales de la primera actividad (*emp_ppal*), clasificación de la población ocupada que trabajan menos de 35 horas más de 35 horas y de 48 horas (*tcco*), clasificación de la población cuenta propia en actividades no calificadas (*cp_anoc*), filas agregadas de la matriz de Husmanns (*mh_fil2*), zona salarial (*zona*) y clasificación de la población ocupada buscador de otro trabajo (*búsqueda*).¹⁰

Es importante resaltar que, en ambos acercamientos, tanto imputación simple como múltiple, se consideraron dos etapas: en primer lugar, se realizó el procedimiento para aquellas observaciones de personas ocupadas con pago que reportaron su ingreso laboral a través de la clasificación en algún rango de salario mínimo y, en segundo lugar, para aquellas observaciones de personas ocupadas con pago que no lo reportaron. Para la imputación simple, las variables de rango de salario mínimo (*p6c*) y zona salarial se incluyeron en la primera parte del proceso; en el caso de los

⁷ Se reconfiguró la variable a tres categorías: 1) ausentes temporales con vínculo laboral, 2) menos de 15 hasta 34 horas y, 3) de 35 a más de 48 horas.

⁸ Se reconfiguró la variable a tres categorías: 1) sin escolaridad o hasta secundaria completa, 2) preparatoria o bachillerato sin tener la educación superior completa y 3) normal, carrera técnica o profesional completa o posgrado (completo o incompleto).

⁹ En estos modelos el número de categorías no ejercen presiones sobre el modelo, por lo que se reconfiguró la variable a cinco categorías: 1) primaria incompleta, 2) primaria completa, 3) secundaria completa, 4) preparatoria o bachillerato sin tener la educación superior completa y, 5) normal, carrera técnica o profesional completa o posgrado (completo o incompleto).

¹⁰ Se agrega la variable a los modelos, ya que ayuda a predecir los ingresos bajos.

modelos de imputación múltiple la zona de salario mínimo se considera en ambas etapas.

Imputación simple

Con el objetivo de recuperar el ingreso laboral, tanto de las observaciones *mvcrsm* como de los *mvsrsm*, se realizó un par de ejercicios empleando el método de imputación simple con medias y medianas, tomando como fuente de información las bases *COE2* y *SDEM* de la ENOE del INEGI. En términos generales, este método contempla tres pasos: en el primero se imputa el ingreso para las observaciones que clasifican su ingreso en algún rango que contemple una cantidad menor a diez salarios mínimos; en el segundo paso, se imputa el ingreso a las observaciones que reportan percibir un ingreso similar al último rango de salarios mínimos, es decir, mayor a diez salarios mínimos; finalmente, en el tercer paso se imputa el ingreso a las observaciones *mvsrsm*.

Es importante aclarar que, el proceso es el mismo tanto para medias como para medianas, la diferencia radica en la medida que se imputa, es decir, en uno se imputa el promedio del ingreso y en otro, la mediana del ingreso, según sea el caso.

A continuación, se detalla cada paso que contempla este método:

1. Para imputar el ingreso laboral a las observaciones *mvcrsm* (menor a diez salarios mínimos), se calcula la media o mediana del ingreso laboral de aquellas observaciones que sí lo reportaron a través de la variable *p6b2* y que se declararon como ocupados con pago, esta medida se obtiene para cada rango de salario mínimo y por zona salarial. Este valor (media o mediana) será el que se impute como ingreso laboral a los *mvcrsm* (menor a diez salarios mínimos).

En este paso existe una excepción, ya que para aquellas observaciones que clasificaron su ingreso laboral en el rango equivalente al valor de un salario mínimo vigente, no se les imputa el promedio o la mediana, sino se les asigna el valor del salario mínimo vigente al momento del levantamiento, el cual se encuentra en la variable *salario*.

2. El ingreso para las observaciones *mvcrsm* en un rango mayor a diez salarios mínimos se imputan en un segundo paso, ya que, a diferencia de los otros seis rangos, este es uno abierto, es decir, no tiene un límite superior. En este paso, primero se generan subcategorías de este grupo de observaciones, las cuales se construyen a partir de todas las combinaciones entre dos variables: la clasificación de la población

ocupada por posición en la ocupación (*pos_ocu*) y la clasificación de empleos formales e informales de la primera actividad (*emp_ppal*). Posteriormente, se calcula una proporción de reporte para cada subcategoría; es decir, de cada subcategoría con ingreso superior a diez salarios mínimos, se calcula el porcentaje de observaciones que sí declara su ingreso (*p6b2*) en términos absolutos. Si este porcentaje es igual o mayor al 65%, se imputa la media o mediana del ingreso reportado por las demás observaciones en ese mismo subgrupo; si es menor, las observaciones se mantendrán como valores faltantes y se imputarán junto con los *mvsrsm*, en el tercer paso.

3. Finalmente, en un tercero y último paso, se imputa el ingreso a las observaciones *mvsrsm*, así como a los valores faltantes que no pudieron ser imputados en el paso anterior. De manera similar a la ronda previa, se generan subcategorías para clasificar a las observaciones. Las categorías en este caso se construyen a partir de todas las combinaciones de cinco variables correlacionadas con el ingreso laboral: nivel de instrucción (*niv_ins*),¹¹ la clasificación de empleos formales e informales (*emp_ppal*), las filas agregadas de la matriz de Hussmanns (*mh_fil2*), la duración de la jornada laboral de la población ocupada (*dur_est*)¹² y la clasificación de la población en actividades no calificadas de cuenta propia (*cp_anoc*).

Para cada una de estas nuevas subcategorías también se calcula una proporción de reporte. Si el porcentaje de observaciones que reportó su ingreso laboral en la variable *p6b2* en una categoría es mayor al 65%, entonces se imputa la media o mediana a los *mvsrsm*. En el caso de que en una subcategoría más del 35% de observaciones declaradas como ocupadas con pago no reporte su ingreso, no se les imputará el ingreso y quedarán como valores perdidos, lo que significa que, para la estimación del porcentaje de la población en pobreza laboral, no se tomarán en cuenta los hogares a los cuales pertenecen dichas observaciones.

Imputación múltiple

La imputación múltiple se realizó considerando tres diferentes modelos, para cada uno de ellos se estimaron cinco imputaciones. La imputación final de cada uno corresponde al promedio del ingreso laboral de las cinco imputaciones. Los modelos utilizados fueron modelo ecuaciones encadenadas con paramiento por medias predictivas, modelo ecuaciones encadenadas con bosques aleatorios y el modelo de imputación múltiple de *Hmisc* (*aregImpute*).

¹¹ Se reconfiguró la variable a tres categorías: 1) sin escolaridad o hasta secundaria completa, 2) preparatoria o bachillerato sin tener la educación superior completa y 3) normal, carrera técnica o profesional completa o posgrado (completo o incompleto).

¹² Se reconfiguró la variable a tres categorías: 1) ausentes temporales con vínculo laboral, 2) menos de 15 hasta 34 horas y, 3) de 35 a más de 48 horas.

El método de imputación múltiple de ecuaciones encadenadas (*Multivariate Imputation by Chained Equations*), MICE por sus siglas en inglés, es un método estadístico para encontrar valores que remplacen los valores perdidos, en este caso de ingreso, pero manteniendo la relación de las variables en la base de datos original, este método se caracteriza por imputar variable por variable un conjunto de datos. Se utilizan dos variantes de este método: paramiento por medias predictivas y bosques aleatorios (Durán, 2019). En caso de no usar la imputación encadenada, la predicción del ingreso dependerá de estas dos variantes.

Imputación múltiple por paramiento por medias predictivas

Paramiento por medias predictivas es una variante del método de *Hot Deck*, en el cual para cada observación con valor perdido del ingreso se le asigna un donador al azar de entre el conjunto de observaciones con valores completos. Este está descrito en Van Buuren, S. (2018), a continuación, se describe de manera general el procedimiento:

1. Se seleccionan las covariantes relevantes para la variable objetivo, es decir, para la variable a imputar.
2. Se realiza una regresión con un modelo de mínimos cuadrados ponderados, con la variable objetivo, en este caso el ingreso laboral, con respecto a las covariables, esta métrica resume la información de las relaciones entre variables de la base de datos.
3. Se estiman los valores predichos de la variable objetivo para las observaciones j con valores faltantes (denotada como \hat{y}_j con $j = 1 \dots n_0$), se agrega aleatoriedad a los coeficientes de regresión.
4. Se estiman los valores predichos de la variable objetivo para las observaciones i con valores completos (denotada como \hat{y}_i con $i = 1 \dots n_1$).
5. Para cada observación con valores faltantes j se selecciona un grupo de donadores Z_j . Este grupo está compuesto por las d observaciones (para este ejercicio se tomó un valor de 5 donadores) con la menor distancia entre la variable predicha y la variable predicha de la observación con dato faltante $|\hat{y}_i - \hat{y}_j|$.
6. Para cada observación con valor faltante j se selecciona un donador de manera aleatoria del grupo de donadores Z_j .
7. El ingreso de este donador i,j se utiliza para imputar el valor faltante de la observación j , $\hat{y}_j = y_{i,j}$.

Imputación múltiple por bosques aleatorios

La imputación múltiple con bosques aleatorios utiliza el algoritmo de clasificación y regresión de Random Forest. Este está descrito en Doove et al. (2014), a continuación, se comenta de manera general el procedimiento para cada iteración:

1. Para cada observación j con valores faltantes ($j = 1 \dots n_0$) se inicia con una primera imputación de la variable objetivo, en este caso el ingreso laboral, y_j^0 obtenida de manera aleatoria del conjunto de observaciones i con valores completos y_i ($i = 1 \dots n_1$).
2. Para cada j se reemplaza y_j^0 con el resultado del ajuste de k árboles aleatorios (para este ejercicio se tomó 10 árboles aleatorios):
 - a) Se obtienen k muestras con el procedimiento de *bootstrapping* de las observaciones i con valores completos y_i .
 - b) Se estima un árbol para cada muestra *bootstrap* k obtenida en el paso anterior, con la selección de un grupo de covariables, los nodos finales (hoja) de cada árbol contienen grupos de observaciones i con valores completos.
 - c) Para cada observación j con valores faltantes se determina el nodo final (hoja) en el que se acomodaría para cada árbol, las observaciones i con valores completos de esos nodos serían los donadores.
 - d) Para cada observación j con valores faltantes, de acuerdo con el nodo final en el que se acomodaría, se agrupan a los donadores de los k árboles para seleccionar el donador final de manera aleatoria $\hat{y}_j = y_{i,j}$.

Imputación múltiple de Hmisc

El algoritmo de imputación múltiple de *Hmisc* (*aregImpute*) es un proceso de paramiento por medias predictivas, pero tiene el objetivo de incorporar en mayor medida la incertidumbre de la información a partir de integrar un proceso de remuestreo por Bootstrap en la estimación. Este algoritmo está descrito por Harrell Jr., F. E. et al. (2024), se comenta de manera general el procedimiento:

1. Para cada observación j con valores faltantes ($j = 1 \dots n_0$) se inicia con una primera imputación de la variable objetivo, en este caso el ingreso laboral, y_j^0 obtenida de manera aleatoria del conjunto de observaciones i con valores completos y_i ($i = 1 \dots n_1$).
2. Se realizan los siguientes pasos para cada iteración:
 - a) Se selecciona una muestra Bootstrap de las observaciones i con valores completos.

- b) Se ajusta un modelo aditivo flexible¹³ para predecir la variable objetivo para cada observación i con valores completos \hat{y}_i .
 - c) Se estima el valor predicho de cada observación j con valores faltantes \hat{y}_j .
 - d) Se realiza la imputación para cada observación j con valores faltantes con la selección de un valor y_i de una observación con valores completos. La selección se obtiene de una distribución multinomial con probabilidades derivadas de pesos basados en las distancias de los valores predichos entre la observación con valor faltante y las observaciones con valores completos $|\hat{y}_i - \hat{y}_j|$.
3. Este proceso se genera para las primeras iteraciones que permiten que las imputaciones converjan a una solución estable (*burnin*). De tal forma que las iteraciones finales generan las imputaciones que se guardarán.

Resultados

Una vez definidos los cinco modelos de imputación, dos de imputación simple (IS. Medias y IS. Medianas) y tres de imputación múltiple (Paramiento por medias predictivas, Bosques aleatorios y HMISC), se estimó la pobreza y el ingreso laboral con cada uno y para la serie del primer trimestre de 2016 al tercer trimestre de 2024 (último periodo disponible al realizar este documento).

En la Gráfica 7, se pueden observar los resultados para cada uno de los métodos, así como el contraste con la pobreza laboral actual publicada por el CONEVAL, cuya metodología fue descrita en el apartado 3 de este documento, *Imputación de ingresos por rangos de salarios mínimos del CONEVAL*. Al respecto, resaltan los siguientes resultados:

En primer lugar, la pobreza laboral con todos los métodos muestra niveles inferiores a la pobreza laboral actual, siendo el método de *IS. Medianas* la que muestra la menor diferencia, la cual oscila entre 2.2 y 3.6 puntos porcentuales menos; en tanto, el método de *HMISC* es el que muestra la mayor brecha, entre 3.9 y 5.6 puntos porcentuales menos.

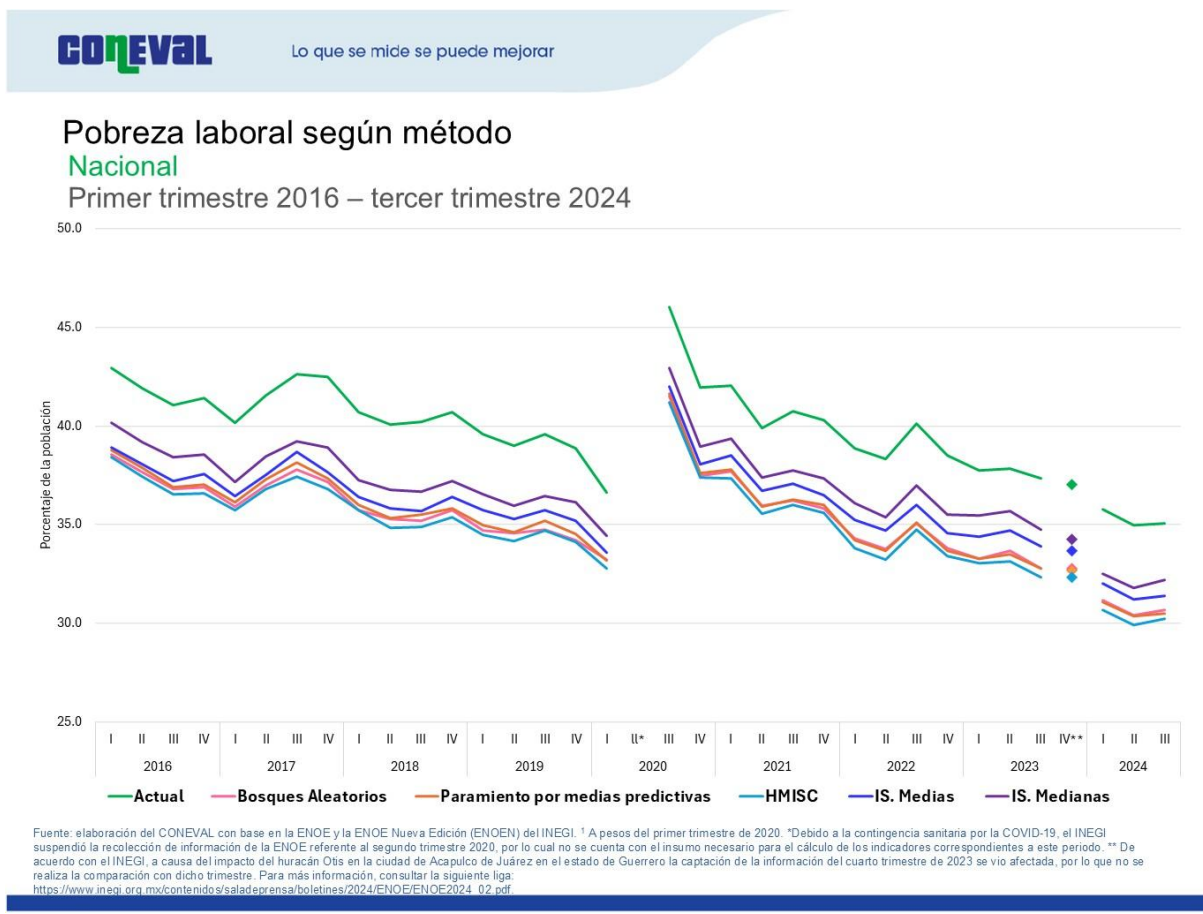
En segundo lugar, en la comparación anual (diferencia de la pobreza laboral con respecto al mismo trimestre de año anterior) todos los métodos de imputación tienen la misma tendencia que la estimación de pobreza laboral actual, a excepción de tres periodos, es decir, varían en la dirección que presenta la pobreza laboral actual. Los tres periodos que muestran diferencias en la dirección de la variación son el primer

¹³ Para este ejercicio no se realizaron transformaciones de la variable objetivo.

trimestre de 2018, el tercer trimestre de 2019 y el tercer trimestre de 2023, periodos en los cuales la variación es muy cercana a cero.

En tercer lugar, se observa que los métodos de imputación múltiple son muy similares entre sí, especialmente, los métodos que provienen de ecuaciones encadenadas, Bosques aleatorios y Paramiento por medias predictivas.

Gráfica 7.



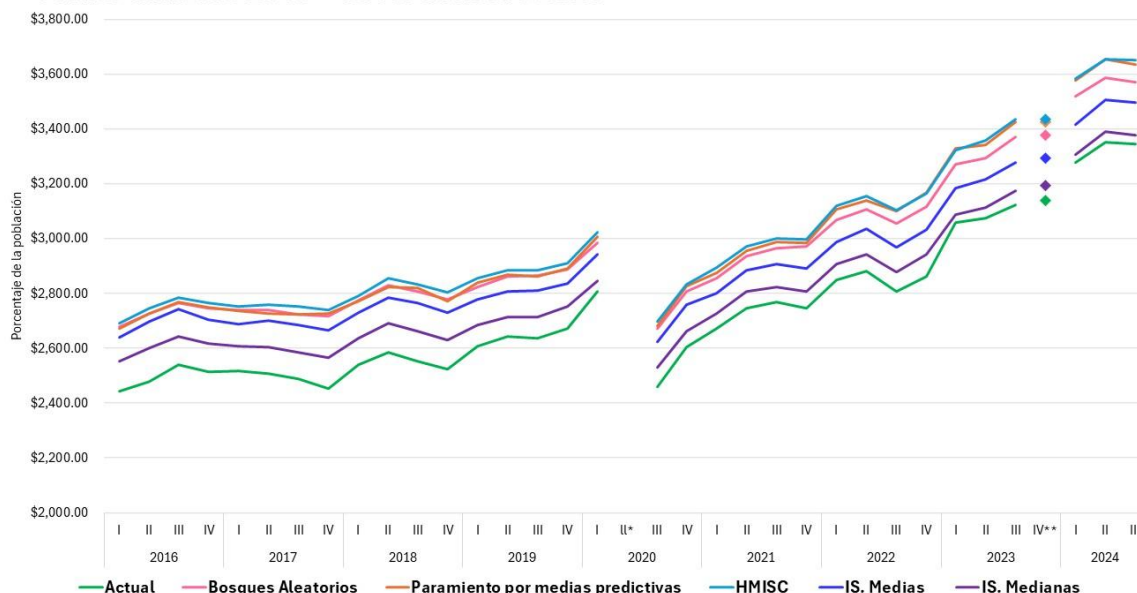
Estos mismos resultados, pero en términos de ingreso laboral, se observan en la Gráfica 8. A excepción de que, en términos de ingreso laboral, la variación de todos los métodos va en la misma dirección que el ingreso laboral actual para todos los periodos.

Gráfica 8.

Ingreso laboral según método

Nacional

Primer trimestre 2016 – tercer trimestre 2024



Fuente: elaboración del CONEVAL con base en la ENOE y la ENOE Nueva Edición (ENOEN) del INEGI. * A pesos del primer trimestre de 2020. ** Debido a la contingencia sanitaria por la COVID-19, el INEGI suspendió la recolección de información de la ENOE referente al segundo trimestre 2020, por lo cual no se cuenta con el insumo necesario para el cálculo de los indicadores correspondientes a este periodo. ** De acuerdo con el INEGI, a causa del impacto del huracán Ots en la ciudad de Acapulco de Juárez en el estado de Guerrero la captación de la información del cuarto trimestre de 2023 se vio afectada, por lo que no se realiza la comparación con dicho trimestre. Para más información, consultar la siguiente liga: https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2024/ENOEN/ENOEN2024_02.pdf

En el proceso de recuperación de los ingresos laborales de las personas ocupadas que no reportaron su ingreso (28,093 observaciones para el tercer trimestre de 2024), la aplicación de estos métodos de imputación permitió recuperar más del 90% de los ingresos laborales no declarados. Sin embargo, a pesar de la alta recuperación a través de los métodos empleados, aún existe un margen de ingresos que no pudo ser recuperado ya que contaban con valores perdidos en alguna de las variables utilizadas para la imputación, dejando un porcentaje de entre el 5% y el 6% en cada uno de los métodos aplicados. La recuperación de los datos, en más del 90%, demuestra la efectividad de los métodos de imputación utilizados. El método de HMISC es el único que, por su metodología y su eficiencia en términos de tiempo, fue utilizado para recuperar el total de las observaciones con ingreso perdido, ya que imputa de igual manera los valores perdidos para las variables insumo, para este ejercicio solo se reporta el ingreso laboral.

Esquema 3.

Número y porcentaje de valores perdidos en el reporte de ingresos de la población ocupada, según método de imputación

ENOE - Tercer trimestre 2024



Fuente: estimaciones del CONEVAL con base en la ENOE del INEGI.

Conclusiones

Los ejercicios de los métodos de imputación presentados permiten recuperar a la mayoría de los valores faltantes de ingreso laboral de la ENOE, lo cual permite a su vez estimar la pobreza laboral, además del porcentaje (valores relativos), en términos de números de personas (valores absolutos), con el fin de ofrecer una herramienta adicional para dimensionar la magnitud de los retos en materia de precariedad laboral y poder adquisitivo del ingreso que existe en el país, así como en las entidades federativas y ciudades.

Estos ejercicios muestran que los resultados para todos los métodos recuperan la misma trayectoria de la variabilidad de la pobreza y el ingreso laboral actuales, aunque con un nivel de pobreza laboral inferior. Lo anterior es un resultado esperado de este ejercicio, debido a que como se presentó anteriormente, las personas que no declaran el ingreso laboral se encuentran principalmente en altos niveles de educación, en trabajos formales, zonas urbanas y son hombres, por lo que, se esperaba que sean personas con ingreso laboral mayor a la media y que, al ser incluidos en la medición de la pobreza laboral, se contabilizan en las personas que no se encuentran en esta situación.

En cuanto a los métodos explorados, estos dan resultados similares en niveles de pobreza laboral, aunque existen algunas diferencias entre ellos. Los métodos de imputación múltiple incorporan en mayor medida la incertidumbre de la información del ingreso laboral, de tal manera que, constituyen una mejor aproximación considerando la variabilidad del conjunto de datos. Sin embargo, la imputación simple representa una aproximación accesible en términos operativos, por lo que, se considera como una alternativa comprensible para el público no especializado.

Estos ejercicios constituyen una propuesta para abordar los datos faltantes de ingreso laboral de la ENOE, la cual tiene como propósito ofrecer información más exhaustiva para estimar pobreza laboral y proveer información en términos absolutos (número de personas) para enriquecer el análisis y diseño de políticas públicas en materia de pobreza y condiciones laborales.

Entre los siguientes pasos, se encuentra un análisis más detallado sobre la medida en que estos métodos recuperan la variabilidad de la información del ingreso laboral y ajustan de mejor manera para identificar a las personas en situación de pobreza laboral.

ANEXO 1

Métodos de imputación de ingresos

En este apartado, se presentan la revisión de los esfuerzos para realizar la imputación de datos faltantes en encuestas de hogares. El tratamiento de los datos faltantes en las bases de datos es variado, para Schafer y Graham (2002) los métodos tradicionales como la eliminación de casos (*case deletion*) y la imputación única (*single imputation*) que han sido ampliamente utilizados, presentan serias limitaciones. Por un lado, la eliminación de casos, aunque simple de implementar, puede inducir sesgos significativos y reducir la eficiencia del análisis al descartar gran parte de la muestra; mientras que, la imputación única, como la sustitución por la media, puede distorsionar la varianza y las correlaciones de los datos, subestimando la incertidumbre y llevando a conclusiones engañosas.

Para mitigar los problemas derivados de la no-respuesta, se han desarrollado diversas técnicas de imputación que buscan completar los datos faltantes y permitir un análisis más robusto y preciso. Entre estas técnicas se encuentran la imputación longitudinal y la imputación transversal. La imputación basada en técnicas longitudinales utiliza información de la misma persona a lo largo del tiempo para imputar valores faltantes. Cuando no se dispone de datos longitudinales, se emplean técnicas transversales como la sustitución lógica, la sustitución de medianas, y la imputación basada en regresión, cada una con sus propias ventajas y limitaciones (Frick & Grabka, 2005).

Schafer y Graham (2002) promueven dos enfoques modernos: la estimación de máxima verosimilitud (ML) y la imputación múltiple (MI, por sus siglas en inglés). Estos métodos, bajo el supuesto de MAR, tienen la capacidad de producir estimaciones no sesgadas y eficientes al utilizar toda la información disponible en los datos observados. La ML integra los valores faltantes en el proceso de estimación, mientras que la MI reemplaza cada dato faltante con múltiples valores simulados, reflejando la incertidumbre inherente a los datos faltantes. Ambos métodos han demostrado ser robustos y adaptables a una amplia gama de situaciones.

Experiencias Internacionales y en México

Organismos internacionales como Statistics Canada y el United States Census Bureau emplean tres métodos para la imputación de valores faltantes: carry-forward, deterministic, y donor (hot-deck) para corregir errores o valores no respondidos en encuestas en hogares. El United States Census Bureau utiliza técnicas como la

imputación relacional, ediciones longitudinales y asignaciones hot-deck en encuestas como la Current Population Survey (CPS) y la Survey of Income and Program Participation (SIPP), con el objetivo de reducir el sesgo de las estimaciones.

Asimismo, existen métodos aplicados por otras instituciones como el Institute for Social and Economic Research de la University of Essex (2018), que utiliza una combinación de imputación hot-deck y regresión en el Panel de Hogares Británico (BHPS). Estas técnicas permiten reducir el sesgo causado por la eliminación de casos con datos faltantes y mantener la variabilidad de los datos originales. Además, en el caso de la Encuesta sobre la Dinámica de los Hogares los Ingresos y el Trabajo (HILDA) en Australia, se emplean métodos como la regresión por vecino más cercano y hot-deck para corregir las diferencias sistemáticas entre los individuos que responden y los que no lo hacen.

En el caso mexicano, existen estudios y análisis que analizan la imputación de ingresos no reportados en la Encuesta Nacional de Ocupación y Empleo (ENOE). Por ejemplo, la Comisión Nacional de los Salarios Mínimos (CONASAMI, 2023) ha implementado la imputación múltiple por ecuaciones encadenadas y el hot-deck aleatorio para intentar minimizar el sesgo en los datos de ingresos laborales. De igual forma, Henderson (2020), utilizó el método hot-deck para medir y dar seguimiento a la desigualdad económica en México a través del Coeficiente de Gini.

Por su parte, Campos (2013) aplica métodos como el pareamiento por puntajes de propensión, hot-deck, y pareamiento por promedios predictivos, y destaca como ventajas la facilidad de interpretación y la eficiencia computacional, mientras que, la necesidad de categorizar variables continuas y la exigencia de recursos computacionales se identifican como desventajas.

Finalmente, Durán (2019), compara diferentes metodologías de imputación aplicadas a los ingresos laborales de la ENOE, y Rodríguez y López (2014), analizan el posible sesgo de ignorar los ingresos faltantes utilizando variantes del método hot-deck. Estos estudios destacan la importancia de elegir métodos adecuados para la imputación de datos, considerando tanto los resultados estadísticos como la aplicabilidad práctica en contextos específicos.

Experiencias internacionales

Respecto al área internacional, se recopilaron principalmente los esfuerzos de organismos oficiales para realizar la imputación de valores faltantes. Entre ellos destaca el Panel de Hogares Británico (BHPS), o bien la Encuesta sobre la Dinámica de los Hogares, los Ingresos y el Trabajo (HILDA) de Australia, ya que estos comparten con la ENOE la característica de dar seguimiento a los hogares e individuos a través del tiempo.

Tabla 1.

Organismos o autor	Instrumento	Objetivo	Métodos de imputación	Comentarios	Fuente
Estadísticas de Canadá (Statistics Canada)	Encuesta de la Fuerza Laboral (LFS)	Imputar valores en los reactivos donde se identifican errores o valores con no respuesta del cuestionario.	Se emplean tres métodos para imputar valores faltantes: <i>carry-forward, deterministic and donor (hot-deck) imputation</i>		https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3701#a2
Buro de Censos de Estados Unidos de América (United States Census Bureau)	Encuesta de Población Actual (CPS) Encuesta de Ingresos y Participación en el Programa (SIPP)	SIPP El objetivo de la imputación es reducir el sesgo de las estimaciones	CPS Se emplean tres métodos para imputar valores faltantes: <i>relational imputation, longitudinal edits and hot deck allocation</i> Para ello ordenan los datos según entidad y UPM dentro de la entidad, esto busca que los datos imputados provengan de la misma región. SIPP En esta encuesta se usan tres diferentes métodos de imputación: <i>model-based imputation; sequential hot-deck imputation; y, cold-deck imputation</i>	En la CPS, la mayoría de los datos imputados respecto al tema laboral se realiza mediante la edición longitudinal, si es el segundo mes o posterior que la persona fue entrevistada, se toma la información previa (en caso de que esté disponible) para imputar. En la SIPP, los procedimientos de imputación utilizados se basan en el supuesto de que los datos perdidos se presentan al azar dentro de subgrupos de la población.	https://www.census.gov/programs-surveys/cps/technical-documentation/methodology/imputation-of-unreported-data-items.html https://www.census.gov/programs-surveys/sipp/methodology/data-editing-and-imputation.html

<p>University of Essex. Institute for Social and Economic Research, (2018)</p>	<p>Panel de Hogares Británico (BHPS)</p>	<p>Reducir el posible sesgo causado por la eliminación de casos con datos faltantes.</p>	<p>Hot- deck-imputación simple se empleó para variables categóricas relacionadas con dinero, por ejemplo, el proxy de la variable de ingreso personal.</p> <p>La imputación de regresión se emplea para variables de cantidad de dinero basadas en una regresión, también se le llama <i>predictive mean matching</i>.</p> <p>La imputación cruzada en panel se emplea en algunas variables en conjunto con las anteriores metodologías, pues es posible que haya información valiosa en diferentes levantamientos.</p>	<p>La imputación de regresión tiene dos ventajas: que se imputa un valor real y que el componente de error aleatorio se añade, de manera que los valores imputados no tienen menos varianza que los reportados.</p>	<p>https://www.understandingsociety.ac.uk/about/british-household-panel-survey</p>
<p>Paulin et al. (2018)</p>	<p>Encuesta de Gastos del Consumidor (CE) Buró de Estadísticas del Trabajo EUA</p>	<p>El propósito es rellenar las respuestas sin valor debido a no respuesta, de tal manera que las inferencias estadísticas puedan ser obtenidas válidamente de los datos.</p>	<p>La imputación se realiza mediante cinco estimaciones diferentes obtenidas mediante regresiones múltiples, donde los coeficientes obtenidos son distorsionados agregando ruido, los cuales son usados para estimar los datos faltantes.</p>	<p>La estimación apropiada de la varianza requiere el uso de las cinco columnas de datos imputados. Por otra parte, si se busca estimar promedios se puede recurrir a la sexta columna que contiene el promedio de las cinco columnas de imputación.</p>	<p>https://www.bls.gov/cex/csxguide.pdf</p> <p>https://www.bls.gov/cex/workshop/2016/microdata/imputation-and-allocation.pdf</p>
<p>Hayes y Watson (2009)</p>	<p>Encuesta sobre la Dinámica de los Hogares, los Ingresos y el Trabajo (HILDA) en Australia</p>	<p>Los individuos que no responden a la entrevista o no proveen respuesta por alguna pregunta muestra diferencias sistemáticas respecto del resto de la muestra, por lo que la imputación busca corregir estas diferencias y mejorar el uso de los datos.</p>	<p>Regresión del vecino más cercano (<i>predictive mean matching</i>), Little and Su, método de arrastre de la población y <i>hot-deck</i></p>	<p>En la regresión con el vecino más cercano se llegan a hacer restricciones adicionales para asegurar que el donador y receptor de algún valor imputado coincidan con características como la edad. Este método es usado cuando los otros no son adecuados. Los receptores con efecto individual cero no son imputados mediante el método Little and Su. Por otra parte, en para emplear este método es necesario que los donadores no tengan un efecto</p>	<p>https://melbourneinstitute.unimelb.edu.au/hilda/publications/hilda-technical-paper-series</p>

				<p>de renglón cero para evitar las divisiones entre cero. En el caso del método de arrastre de la población, este se emplea para la identificación de cantidades cero o no cero, sin embargo, para imputar la cantidad no cero se utilizan otros métodos.</p>	
--	--	--	--	---	--

Propuestas en México

El aumento en la proporción de los valores perdidos en el ingreso reportado en la ENOE ha derivado en diversos estudios y publicaciones enfocados en el caso mexicano, por lo que en esta sección se presentan algunos de ellos, con el fin de recopilar aquellos que hicieron uso de uno o varios métodos de imputación, los cuales también se revisaron y analizaron para la selección del método de imputación que el CONEVAL ha decidido aplicar para la estimación del porcentaje de la población en pobreza laboral.

Tabla 2.

Autor(es), año	Título	Objetivo	Métodos de imputación	Comentarios	Liga de consulta
Comisión Nacional de los Salarios Mínimos (CONASAMI), 2023	Nota metodológica: Imputación de ingresos laborales no reportados en la ENOE	Minimizar el sesgo de estimación en las cifras de ingresos laborales en las publicaciones de la CONASAMI, cuya fuente sea la ENOE, a través de la aplicación del método de imputación múltiple por ecuaciones encadenadas por pareamiento por medias predictivas (PMM)	Imputación múltiple por ecuaciones encadenadas por pareamiento por medias predictivas (PMM) y Hot-deck (HD) aleatorio	Si bien se aplican los dos métodos, a través de la aplicación de la raíz del error cuadrático medio (RMSE) se definió que el PMM era el que tenía mejor desempeño. En el texto se menciona que el método seleccionado (PMM) se basa en el supuesto de que la ausencia del ingreso no depende del ingreso en sí, sino de una serie de características observables, es decir, está relacionado con el patrón de los datos observados.	https://www.gob.mx/cms/uploads/attachment/file/806698/Imputacion_de_ingresos_no_reportados_en_la_ENOE.pdf
Max Henderson, 2020	Seguimiento de la desigualdad económica en México	Medir y dar seguimiento regular a la desigualdad en México, a través de las series de datos sobre desigualdad de 2004 a 2019 para calcular el Coeficiente de Gini a partir del ingreso laboral	Hot-deck (no paramétrico)	No se mencionan ventajas o desventajas	https://sobremexico-revista.ibero.mx/index.php/Revista_Sobre_Mexico/article/view/88/41
Raymundo M. Campos-Vázquez, 2013	Efectos de los ingresos no reportados en el	Calcular ingresos para aquellas personas que no declaran ingresos, de tal forma	1) Pareamiento por puntajes de propensión;	De acuerdo con el autor, algunas de las ventajas del Pareamiento por puntajes de propensión son:	https://ensayos.uan.mx/index.php/ensa

	<p>nivel y tendencia de la pobreza laboral en México</p>	<p>que, se puedan analizar las tendencias de pobreza y desigualdad, si no existiera el problema de los ingresos inválidos</p>	<p>2) Hot-deck; 3) Por grupos con aleatoriedad; y 4) Pareamiento por promedios predictivos</p>	<ul style="list-style-type: none"> - El ingreso imputado es un valor observado (en caso de utilizar el método del vecino más cercano) - Fácil interpretación y entendimiento del método <p>Por el contrario, algunas de las desventajas son:</p> <ul style="list-style-type: none"> - Método intensivo, computacionalmente - En STATA, no existe una rutina para realizar la imputación - No es posible la extensión a imputación múltiple <p>Algunas ventajas que comenta para el método Hot-deck:</p> <ul style="list-style-type: none"> - Los valores imputados son valores observados - Fácilmente aplicado a imputación única o múltiple - Eficiente, en término de tiempo computacional - El método se utiliza con el mismo fin en otros países - Cuenta con una rutina en STATA, lo que evita errores de programación y lo hace de fácil aplicación <p>Una de las desventajas que el autor menciona es que, al igual que la mediana por grupos, es que cada grupo tiene que ser definido categóricamente, es decir que si se tienen variables continuas es necesario categorizar en grupos, lo que podría abrir paso a la pérdida de información.</p> <p>Respecto al método por grupos con aleatoriedad, menciona las siguientes ventajas:</p> <ul style="list-style-type: none"> - Sencilla aplicación en programación y su rápida ejecución 	<p>yos/article/view/58/44</p>
--	--	---	--	---	---

				<ul style="list-style-type: none"> - Cuenta con extensión en imputación única o múltiple Por otro lado, menciona las siguientes desventajas: <ul style="list-style-type: none"> - Se asume siempre que la mediana y la desviación observadas dentro de cada grupo son las correctas - Si la distribución de cada grupo no sigue una normal, se estaría cambiando la distribución verdadera de los datos Finalmente, respecto al pareamiento por promedios predictivos, el autor menciona las siguientes ventajas: <ul style="list-style-type: none"> - Cuenta con robustez estadística - Valores imputados de valores observados - Sencilla expansión a imputación múltiple - Fácil implementación en STATA Mientras que sólo menciona una desventaja que consiste en que, si la distribución de cada grupo no sigue una normal, se estaría cambiando la distribución verdadera de los datos 	
Benito Durán Romo, 2019	Comparación de metodologías de imputación aplicadas a ingresos laborales de la ENOE	Mostrar los resultados de un ejercicio comparativo de algoritmos y metodologías de imputación de ingresos laborales de la ENOE para ponerlas a consideración de los usuarios de la encuesta y valorar su posible adopción como solución a los ingresos laborales faltantes.	<ul style="list-style-type: none"> - Imputación simple: hot-deck aleatorio y missForest - Imputación múltiple: pareamiento por medias predictivas, MICE (RF), Amelia II y Rf2e 	El autor menciona que los expertos han mostrado preferencia por las técnicas de imputación múltiple, por lo que, de acuerdo con los resultados de diversas pruebas (), resultaría buena opción elegir entre MICE (pmm), Rf2e, Hmisc y MICE (RF), aunque MICE (pmm) es la técnica que arrojó los mejores resultados en dichas pruebas	https://rde.inegi.org.mx/wp-content/uploads/2019/09/RDE29_art01_2c.pdf
Eduardo Rodríguez-Oreggia y Bruno López Videla, 2014	Imputación de ingresos laborales: una aplicación con	Analizar el posible sesgo que existe de ignorar los ingresos faltantes en la ENOE al medir los efectos de diversas	Hot-deck dos variantes: mediante asignación aleatoria y mediante	Ventajas: <ul style="list-style-type: none"> - Ambos métodos son empleados por el Bureau of Labor Statistics de Estados 	https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2448-



Consejo Nacional de Evaluación
de la Política de Desarrollo Social

	encuestas de empleo en México	variables sobre el ingreso y pobreza laboral, una vez imputados los ingresos mediante el método Hot-deck	una función de distancia de Mahalanobis	Unidos, para la imputación de salarios en la Current Population Survey (CPS) - Mantienen la misma distribución de características para cada celda k	718X2015000100117
--	-------------------------------	--	---	--	-----------------------------------

ANEXO 2

Variables exploradas

Se utilizó una lista de variables precodificadas que ofrece la ENOE referente a las características socioeconómicas y de ocupación de las personas ocupadas que reportan y no reportan ingresos.

Tabla 3.

Variable	Etiqueta	Valores
sex	sexo	1 Hombre 2 Mujer
edad7c	Clasificación de la población de 15 años y más: grupo de edad 7	1 De 15 a 19 años 2 De 20 a 29 años 3 De 30 a 39 años 4 De 40 a 49 años 5 De 50 a 59 años 6 De 60 años y más 7 Edad no especificado
niv_ins	Nivel de instrucción	1 Primaria incompleta 2 Primaria completa 3 Secundaria completa 4 Medio superior y superior 5 No especificado
pos_ocu	Población ocupada por posición en la ocupación	1 Trabajadores subordinados y remunerados 2 Empleadores 3 Trabajadores por cuenta propia 4 Trabajadores sin pago 5 No especificado
c_ocu11c	Población ocupada por condición de ocupación	1 Profesionales, técnicos y trabajadores del arte 2 Trabajadores de la educación 3 Funcionarios y directivos 4 Oficinista

		<p>5 Trabajadores industriales artesanos y ayudantes 6 Comerciantes 7 Operadores de transporte 8 Trabajadores en servicios personales 9 Trabajadores en protección y vigilancia 10 Trabajadores agropecuarios 11 No especificado</p>
tue1	Población ocupada por tipo de unidad económica - Totales	<p>1 Empresas y negocios 2 Instituciones 3 Sector de los hogares 4 Situaciones de carácter especial y no especificadas</p>
dur_est	Población ocupada por duración de la jornada	<p>1 Ausentes temporales con vínculo laboral 2 Menos de 15 horas 3 De 15 a 34 horas 4 De 35 a 48 horas 5 Más de 48 horas 6 No especificado</p>
emp_ppal	Empleos formales e informales de la primera actividad	<p>1 Empleo informal 2 Empleo formal</p>
tcco	Población ocupada que trabajan menos de 35 horas, más de 35 horas y de 48 horas	<p>1 OH35RM (Ocupados que trabajan menos de 35 horas por razones de mercado) 2 OH35SM (Ocupados que trabajan de 35 horas o más y ganan hasta un salario mínimo) 3 OH48SM (Ocupados que trabajan más de 48 horas con ingresos de más de 1 hasta 2 salario mínimo)</p>
cp_anoc	Población cuenta propia en actividades no calificadas	<p>1 Cuenta propia en actividades no calificadas (Para promedio y medias)</p>
mh_fil2	Filas agregadas de la matriz	<p>1 Sector informal 2 Trabajo doméstico remunerado 3 Empresas, Gobierno e Instituciones 4 Ámbito agropecuario</p>
busqueda	Clasificación de la población ocupada	<p>1 Con condición de búsqueda de un trabajo adicional para tener dos empleos 2 Sin condición de búsqueda de un trabajo adicional</p>

Bibliografía

Campos, R. (2013, 2 de noviembre). *Efectos de los ingresos no reportados en el nivel y tendencia de la pobreza laboral en México*. Disponible en: <https://ensayos.uanl.mx/index.php/ensayos/article/view/58/44>

CONASAMI. (2023, 28 de febrero). *Nota metodológica. Imputación de ingresos laborales no reportados en la ENOE*. Disponible en: https://www.gob.mx/cms/uploads/attachment/file/806698/Imputacion_de_ingresos_no_reportados_en_la_ENOE.pdf

Doove, L.L., van Buuren, S., Dusseldorp, E. (2014), Recursive partitioning for missing data imputation in the presence of interaction Effects. *Computational Statistics & Data Analysis*, 72, 92-104. Disponible en: <https://stefvanbuuren.name/publications/2014%20Recursive%20partitioning%20-%20CSDA.pdf>

Durán Romo, B. (2019, septiembre-diciembre). *Comparación de metodologías de imputación aplicadas a ingresos laborales de la ENOE*. *RDE*, 29. Disponible en: https://rde.inegi.org.mx/wp-content/uploads/2019/09/RDE29_art01_2c.pdf

Frick, J. R., & Grabka, M. M. (2005). *Item nonresponse on income questions in panel surveys: Incidence, imputation, and the impact on inequality and mobility*. *Allgemeines Statistisches Archiv*, 89(1), 49-61. <https://doi.org/10.1007/s101820500191>

Harrell Jr., F. E., et al. (2024). *Hmisc: Harrell Miscellaneous* (versión del paquete R 4.8-0). Disponible en: <https://cran.r-project.org/web/packages/Hmisc/Hmisc.pdf>

Hayes, & Watson. (2009). *HILDA imputation methods*. Disponible en: <https://melbourneinstitute.unimelb.edu.au/hilda/publications/hilda-technical-paper-series>

Henderson, M. (2020, 30 de septiembre). *Seguimiento de la desigualdad económica en México*. Disponible en: https://sobremexico-revista.iberomexico.mx/index.php/Revista_Sobre_Mexico/article/view/88/41

INEGI. (2024). *Encuesta Nacional de Ocupación y Empleo (ENOE), población de 15 años y más de edad*. Disponible en: <https://www.inegi.org.mx/programas/enoe/15ymas/>

Luján Salazar, J. de J. (2009, agosto). *Tesis. Imputación del ingreso en la ENOE*. Disponible en: <https://cimat.repositorioinstitucional.mx/jspui/bitstream/1008/200/2/TE%20286.pdf>

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2ª ed.). Wiley.

Max Henderson. (2020, 30 de septiembre). *Seguimiento de la desigualdad económica en México*. Disponible en: https://sobremexico-revista.iberomexico.mx/index.php/Revista_Sobre_Mexico/article/view/88/41

Paulin, et al. (2018). *User's guide to income imputation in the Consumer Expenditure Survey (CE)*. Bureau of Labor Statistics. Disponible en: <https://www.bls.gov/cex/csxguide.pdf>

Rodríguez-Oreggia, E., & López Videla, B. (2014, septiembre). *Imputación de ingresos laborales: Una aplicación con encuestas de empleo en México*. El Trimestre Económico. Disponible en: <https://www.eltrimestreeconomico.com.mx/index.php/te/article/view/142/142>

Schafer, J. L., & Graham, J. W. (2002). *Missing data: Our view of the state of the art*. Psychological Methods, 7(2), 147-177. Disponible en: <https://doi.org/10.1037/1082-989X.7.2.147>

Schräpler, J.-P. (2004). *Respondent behavior in panel studies: A case study for income nonresponse by means of the German Socio-Economic Panel (SOEP)*. Sociological Methods & Research, 33(1), 118-156. Disponible en: <https://doi.org/10.1177/0049124103262689>

Si, Y., & Reiter, J. P. (2013). *Nonparametric Bayesian multiple imputation for incomplete categorical variables in large-scale assessment surveys*. Journal of Educational and Behavioral Statistics, 38(5), 499-521. Disponible en: <https://doi.org/10.3102/1076998613480394>

Statistics Canada. (2023). *Labour Force Survey (LFS) Data sources and methodology*. Disponible en: <https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3701#a2>

United States Census Bureau. (2023). *Current Population Survey (CPS) Imputation of unreported data items*. Disponible en: <https://www.census.gov/programs-surveys/cps/technical-documentation/methodology/imputation-of-unreported-data-items.html>

United States Census Bureau. (2023). *Survey of Income and Program Participation (SIPP) Data editing and imputation*. Disponible en: <https://www.census.gov/programs-surveys/sipp/methodology/data-editing-and-imputation.html>

University of Essex, Institute for Social and Economic Research. (2023). *User manual of the British Household Panel Survey (BHPS)*. Disponible en: <https://www.understandingsociety.ac.uk/about/british-household-panel-survey>

Van Buuren, S. (2018). *Flexible imputation of missing data* (2ª ed.). Chapman & Hall/CRC.