**Aclaraciones metodológicas sobre el armado de la base de datos del proyecto “Precariedad Mundial”**

El proyecto de “Precariedad Mundial” tiene como objetivo aportar argumentos y evidencias empíricas sobre la incidencia de la precariedad laboral a lo largo del mundo. Este documento está orientado a detallar las decisiones metodológicas tomadas a la hora de construir una base de datos internacional a partir de una homogeneización de un conjunto de variables presentes en las encuestas hogares de distintos países. Antes de avanzar con lo estrictamente metodológico resultan útiles dos aclaraciones.

La primera es que la noción de “empleo precario” no es unívoca en la literatura. No sólo su delimitación y cuantificación suele ser diversa en los trabajos académicos, sino que las encuestas a hogares de distintos países presentan distintas preguntas que se asocian a la posibilidad de captar aquellas expresiones de la precariedad laboral que resultan más frecuentes en cada país. En nuestra base de datos recabamos 4 expresiones de la precariedad aun cuando estas no pueden evaluarse en todos los países: **trabajo part-time involuntario, el no registro de la relación laboral, el trabajo de duración determinada y la falta de aportes a la seguridad social**. A su vez, consideramos que la dimensión de los **ingresos percibidos** en una ocupacióntambién es expresiva de la calidad del empleo, por lo cual, incluimos dicha variable en nuestra base.

La segunda es que la selección de variables adicionales incluidas en la base no es ajena a los fundamentos teóricos que sustentan nuestra línea de investigación. En los artículos citados en el repositorio se brindan los argumentos por los cuales nos enfocamos centralmente en dos variables: **la** **calificación del puesto y el tamaño de los establecimientos**. Más allá de eso, la base provee un conjunto adicional de variables que pueden ser de interés en general, tales como el sexo, edad, nivel educativo, entre otras.

Hechas estas aclaraciones generales, a continuación se detallan los procedimientos que fueron realizados para la construcción del dataset. En primer lugar, el siguiente cuadro especifica las encuestas utilizadas para cada país, su periodicidad y el año para el cual se procesó la información.

# Cuadro 1. Encuestas utilizadas por país y periodicidad

|  |  |  |  |
| --- | --- | --- | --- |
| **País** | **Encuesta** | **Periodicidad** | **Año utilizado** |
| Argentina | Encuesta Permanente de Hogares | Trimestral | 2019 |
| Bolivia | Encuesta Continua de Empleo | Trimestral | 2019 |
| Brasil | Pesquisa Nacional por Amostra de Domicílios Contínua - PNAD Contínua | Trimestral | 2019 |
| Chile | Encuesta Nacional de Empleo (ENE) – módulo Encuesta Suplementaria de Ingresos (ESI) | Anual | 2019 |
| China | Chinese Household Income Proyect (CHIP) | Anual | 2018 |
| Colombia | Gran Encuesta Integrada de Hogares - GEIH | Mensual | 2019 |
| Costa Rica | Encuesta Nacional de Hogares | Anual | 2019 |
| Ecuador | Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU) | Mensual | 2019 |
| El Salvador | Encuesta Nacional de Hogares | Anual | 2019 |
| Estados Unidos | Current Population Survey | Mensual | 2018 |
| \*Europa | Eurostat Labour Force Survey (LFS) | Anual | 2018 |
| Guatemala | Encuesta Nacional de Empleo e Ingresos | Semestral | 2019 |
| Mexico | Encuesta Nacional de Ocupación y Empleo | Trimestral | 2019 |
| Paraguay | Encuesta Permanente de Hogares Continua | Anual | 2019 |
| Perú | Encuesta Nacional de Hogares | Trimestral | 2019 |
| Uruguay | Encuesta Continua de Hogares | Anual | 2019 |

\*La Eurostat LFS condensa la información de distintos países: En nuestra base se encuentran todos los países con información para 2018: Grecia, Polonia, Italia, Portugal, España, Austria, Reino Unido, Países Bajos, Noruega, Francia, Dinamarca, Suecia y Alemania.

En los casos de países donde existe más de una encuesta a hogares con información sobre el mercado laboral, se privilegió aquella que presentara las variables asociadas al tamaño de los establecimientos y a la calificación del puesto. Por otra parte, en los casos donde la periodicidad de la encuesta no es anual, los datos presentados en el dataset consisten en un *pool* de las encuestas trimestrales o mensuales, conservando en la variable PERIODO la información sobre la correspondencia de cada caso a un trimestre o mes específico. Cabe aclarar que no se realizó un procedimiento de eliminación de registros de hogares que puedan aparecer en distintos períodos dados los esquemas de rotación de las muestras.

Dados el propósito de asegurar la mayor comparabilidad en las estimaciones y considerando que en algunos países las encuestas a hogares contemplan las áreas rurales y en otros no, se realizó un filtro por el cual todos los casos presentados en la base corresponden únicamente a **áreas urbanas.** Por otra parte, acorde a los propósitos de nuestra comparación, la base solo presenta información para las **personas ocupadas**, mientras que la información sobre calidad del empleo e ingresos refiere a la **ocupación principal** de dichas personas.

Cabe aclarar que, si bien en los artículos de investigación incluimos un filtro adicional para conservar solo el universo constituido por los cuentapropistas y por los asalariados del sector privado excluyendo el servicio doméstico, en la presentación de la base de datos optamos por no realizar este filtro de categorías ocupacionales y sectores a los fines de otorgar mayor flexibilidad a quien quiera incluir en su análisis a los restantes casos. De esta forma, incluimos una variable de **categoría ocupacional** y de **sector**, con una reclasificación hacia las principales categorías que pueden distinguirse en cada una de las encuestas**.**

# Cuadro 2. Recorte del diccionario de la base. Tomado de archivo “Metadata.xlsx”

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **variable** | **tipo** | **descripción** | **valores** | **referencia** |
| CATOCUP | character | Categoría ocupacional | Patron | Patrón |
| Asalariado | Asalariado |
| Cuenta Propia | Cuenta Propia |
| Resto | Resto |
| SECTOR | character | Sector | Pub | Publico |
| Priv | Privado |
| SD | Servicio Doméstico |

El detalle sobre el conjunto de variables que integran el dataset y sus correspondientes categorías puede observarse en el archivo “Metadata.xlsx” que integra el directorio principal del repositorio. Una primera pestaña de dicho archivo (llamada Diccionario) muestra como aparecen nomencladas cada una de las variables en el dataset, el tipo de datos que contienen, su descripción, los valores (tal cual figuran en la base) y la referencia sobre qué indican dichos valores.

Por otra parte, el procedimiento de construcción de estas variables a partir de la homogeneización de la información provista por las encuestas de cada país puede observarse en una segunda pestaña de dicho archivo (llamada Homogeneización). Allí se muestra, para cada país, el detalle de las preguntas y categorías originales provistas en la encuesta y su reclasificación hacia las variables de nuestra base. A continuación se muestra un ejemplo para dos países del procedimiento realizado para homogeneizar la variable de tamaño de establecimiento.

# Cuadro 3. Ejemplo de homogeneización para variable de tamaño del establecimiento. Tomado de archivo “Metadata.xlsx”

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **País** | **Año** | **Encuesta** | **Variable original** | **Definición** | **Valor** | **Etiqueta** | **Variable dataframe** | **Reclasificación** |
| China | 2018 | Chinese Household Income Proyect (CHIP) | C08 | How many employees were at your work unit of this job? | 1 | Hasta 8 | TAMA | Pequeño |
| 2 | 9 a 50 | TAMA | Mediano |
| 3 | 51 a 100 | TAMA | Grande |
| 4 | 101 a 250 | TAMA | Grande |
| 5 | 251 a 500 | TAMA | Grande |
| 6 | 501 a 1000 | TAMA | Grande |
| Bolivia | 2019 | Encuesta Continua de Empleo | s2\_26 | ¿Cuántas personas trabajan en la empresa, institución o lugar donde trabaja, incluido usted? | numérico | 1-10 | TAMA | Pequeño |
| 11 a 49 | TAMA | Mediano |
| > 50 | TAMA | Grande |

Tras observar las categorías disponibles a lo largo de todas las encuestas nuestra reclasificación propone tres categorías idealmente construidas de la siguiente manera: pequeño (hasta 10 personas), mediano (11 a 49 personas) y grande (50 o más personas). Lógicamente, no todas las encuestas presentan el mismo punto de corte entre categorías, por ende, como se muestra en el ejemplo para el caso de China, la reclasificación no es exacta.

# Aclaraciones específicas por país

**Labour Force Survey (Europa)**

La versión de la LFS con la que se trabajó en este repositorio no contiene información sobre ingresos laborales, estando solamente disponible la información sobre el decil de ingreso al que pertenece la persona encuestada. Por ello se decidió complementar la información de dicha base con información sobre ingresos de la Structure of Earnings Survey también elaborada por Eurostat. Vale la pena remarcar que por diversos motivos dicha imputación tiene un carácter exploratorio y sus datos no pueden considerarse estadísticamente representativos de los ingresos laborales de la población analizada por la LFS.

**Alemania:** se utilizaron datos del año 2017 debido a problemas con los datos de la variable SIZE en la base original para el año 2018.