

Proyecto:

Efectos de la desigualdad socioeconómica en la percepción de la democracia en Latinoamérica: periodo 1995-2023

Ficha Metodológica:

Latinobarómetro

Versión 001

Versión	Causa	Responsable	Fecha
001	Versión Inicial	Patricio Mendoza	23/11/2025

Contents

1 Metodología para Latinobarómetro	3
1.1 Descripción general de la fuente	3
1.1.1 Descarga, conversión y organización de los datos	3
1.1.2 Limpieza y normalización de variables	3
1.1.3 Cálculo de indicadores	5
1.1.4 Desagregación geográfica y temporal	6
2 Inventario de indicadores	6
2.1 Apoyo a la democracia (<i>democ_supp</i>)	6
2.2 Indiferencia o apoyo a alternativas autoritarias	7
2.3 Satisfacción con la democracia (<i>democ_satis</i>)	7
2.4 Ubicación ideológica izquierda–centro–derecha (<i>left_right_scale</i>)	7
2.5 Evaluación de la situación económica del país (<i>econ_situation</i>)	8
2.6 Evaluación de la situación económica personal (<i>resp_economic_perception</i>)	8
2.7 Preocupación por perder el empleo (<i>job_concern</i>)	8
2.8 Tenencia de bienes y servicios en el hogar (<i>goods_car</i> , <i>goods_hot_water</i> , <i>goods_sewage</i> , <i>goods_wash_mach</i>)	8
2.9 Confianza en instituciones	9
2.10 Otras variables normalizadas de contexto	9
3 Limitaciones y supuestos	9

1 Metodología para Latinobarómetro

1.1 Descripción general de la fuente

Latinobarómetro es una encuesta de opinión pública aplicada de forma periódica en países de América Latina (y España en algunos años). Proporciona información sobre actitud frente a la democracia, instituciones, economía, bienes del hogar y otros temas de interés social y político.

En este proyecto, los microdatos anuales se integran en una base unificada en ClickHouse ([indicadores.latinobarometro](#)), sobre la cual se calculan indicadores agregados por país, región dentro de país y ciudad, para todos los años disponibles.

1.1.1 Descarga, conversión y organización de los datos

- **Descarga de microdatos:** los datos de Latinobarómetro se obtienen desde los repositorios oficiales del proyecto (o mirrors institucionales) en formato `.zip`, que a su vez contienen archivos `.dta` de Stata.
- **Organización de archivos:**
 - Para cada año de la encuesta se mantiene una carpeta de datos crudos (`raw`), con los `.dta` originales.
 - El pipeline de automatización se encarga de extraer cada archivo `.dta`, convertirlo a CSV y almacenarlo en un directorio `processed` con un nombre homogéneo (por ejemplo, `latinobarometro_YYYY.csv`).
- **Conversión `.dta` → CSV:** la conversión se realiza controlando:
 - Conservación de etiquetas de valor cuando es posible (para documentación externa).
 - Inclusión de todas las columnas originales, incluso si en fases posteriores se selecciona un subconjunto.
 - Codificación de caracteres en UTF-8 para evitar problemas en la base de datos.
- **Carga en ClickHouse:** los CSV resultantes se cargan en la tabla `indicadores.latinobarometro`, definiendo las columnas clave:
 - `research_year`: año de la encuesta.
 - `resp_country`, `research_region`, `research_city`: identificadores numéricos de país, región y ciudad.
 - Variables normalizadas de interés: `democ_supp`, `democ_satis`, `left_right_scale`, `econ_situation`, variables de bienes del hogar, confianza institucional, etc.
 - Se generan además columnas alias `country_name` e `iso3` a partir del código numérico de país (`resp_country`), utilizando transformaciones codificadas en el esquema SQL.

1.1.2 Limpieza y normalización de variables

Un desafío central del uso de Latinobarómetro es que, aunque las preguntas se mantienen conceptualmente estables a lo largo del tiempo, sus *códigos de variable* y, en ocasiones, las categorías de respuesta cambian de un año a otro. Por ejemplo, la pregunta sobre apoyo a la democracia puede aparecer como `p1`, `p3` o `s1` según el año, y las categorías pueden reordenarse o renombrarse, aun cuando el contenido sustantivo sea equivalente.

Para poder construir series de tiempo e indicadores comparables, se implementó un proceso explícito de **normalización de preguntas y variables**, documentado en el diccionario:

`latinobarometro_glossary_candidates_BM.csv`

Este diccionario registra, para cada combinación de año y pregunta original, al menos cuatro elementos:

- **Año de la encuesta (year).**
- **Nombre de la variable original** tal como aparece en el archivo de Latinobarómetro (`question`).
- **Etiqueta o descripción de la pregunta (label)**, que permite verificar que el significado sustantivo sea el mismo.
- **Nombre de la variable normalizada (normalization)**, que agrupa bajo un identificador común preguntas equivalentes en distintos años (por ejemplo, `democ_supp`, `democ_satis`, `left_right_scale`, `econ_situation`, `goods_car`, entre otras).

En la etapa de limpieza se aplican los siguientes pasos generales:

1. **Unificación de nombres de variables.** Para cada año, las variables originales se renombraron utilizando el diccionario, de modo que todas las mediciones conceptualmente equivalentes comparten el mismo nombre normalizado (p. ej. todas las versiones de la pregunta “La democracia es preferible a cualquier otra forma de gobierno” se renombraron como `democ_supp`).
2. **Revisión manual de equivalencias.** El diccionario no se construyó de manera puramente automática: se revisaron etiqueta por etiqueta para asegurar que las preguntas asociadas a un mismo nombre normalizado fuesen realmente comparables en contenido, redacción y escala de respuesta. Cuando existían cambios sustantivos en la pregunta o en las alternativas, se decidió no agruparlas o se documentaron las diferencias.
3. **Normalización de categorías de respuesta.** Para cada variable normalizada se definieron categorías comunes (por ejemplo, *much/a/algo/poca/ninguna confianza*, *muy buena/buena/mala/muy mala*, *izquierda/centro/derecha*). Las categorías de cada año fueron recodificadas a estos esquemas comunes, tratando explícitamente valores como “*no sabe*”, “*no contesta*” o *faltantes* como no válidos y excluyéndolos de los denominadores de los indicadores.
4. **Control de calidad y consistencia.** Se generaron tablas de frecuencia por año y país para cada variable normalizada, comparando distribuciones antes y después de la recodificación. Desviaciones inesperadas (por ejemplo, cambios bruscos en la proporción de respuestas válidas o en la distribución de categorías) se utilizaron para detectar problemas de codificación o mapeos incorrectos en el diccionario.

El resultado de este proceso es una base armonizada en la que cada indicador se construye siempre a partir de los mismos nombres de variables normalizadas. El diccionario `latinobarometro_glossary_candidates_BM.csv` es, por lo tanto, un insumo central: garantiza la **trazabilidad** entre el archivo original de Latinobarómetro y las variables analíticas utilizadas en las vistas e indicadores finales.

1.1.3 Cálculo de indicadores

1.1.3.1. Tabla de sumas latinobarometro_sums

- La tabla `latinobarometro_sums` almacena sumas por combinación (`level`, `anio`, `periodo`, `area`, `geo`), donde:
 - `level` ∈ {`pais`, `region`, `ciudad`, `pais_total`}.
 - `anio` es el año de aplicación de la encuesta.
 - `periodo=0` y `area=0` son campos de compatibilidad con el resto del esquema.
 - `geo` identifica el país, la región o la ciudad, según corresponda.
- La vista materializada `mv_latinobarometro_all_levels` realiza:
 - Cálculo de banderas de validez para cada indicador (por ejemplo, `democ_supp_v`, `democ_satis_v`, `lr_v`, `econ_v`, `conf_congress_v`, etc.).
 - Cálculo de numeradores para cada categoría de interés, entre ellos:
 - * $s_{democ_support_dem} = \sum w \cdot \mathbb{1}(democ_supp = 1)$.
 - * $s_{democ_support_auth} = \sum w \cdot \mathbb{1}(democ_supp = 2)$.
 - * $s_{democ_satis} = \sum w \cdot \mathbb{1}(democ_satis = 1)$.
 - * $s_{left_lean} = \sum w \cdot \mathbb{1}(0 \leq left_right_scale \leq 4)$.
 - * $s_{right_lean} = \sum w \cdot \mathbb{1}(5 \leq left_right_scale \leq 10)$.
 - * $s_{econ_good} = \sum w \cdot \mathbb{1}(econ_situation \in \{1, 2\})$.
 - * $s_{econ_bad} = \sum w \cdot \mathbb{1}(econ_situation \in \{4, 5\})$.
 - * $s_{goods_wash_mach}$, s_{goods_car} , s_{goods_sewage} , $s_{goods_hot_water}$: bienes presentes.
 - * $s_{conf_*_high}$ y $s_{conf_*_low}$: confianza alta ($\in \{1, 2\}$) y baja ($\in \{3, 4\}$) en Congreso, Poder Judicial, Iglesia, Policía, Fuerzas Armadas y partidos políticos.
 - Cálculo de denominadores de casos válidos para cada indicador (por ejemplo, `s_valid_democ_supp`, `s_valid_democ_satis`, `s_valid_lr`, `s_valid_econ_sit`, etc.).

1.1.3.2. Vistas de indicadores en porcentaje

- A partir de `latinobarometro_sums` se generan vistas que calculan porcentajes por nivel geográfico:
 - `indicadores_latino_pais`: indicadores agregados por país y año.
 - `indicadores_latino_region`: indicadores por región dentro de país y año.
 - `indicadores_latino_ciudad`: indicadores por ciudad dentro de país y año.
- Ejemplos de fórmulas en la vista por país:

$$tasa_apoyo_democracia = 100 \cdot \frac{\sum s_{democ_support_dem}}{\sum s_{valid_democ_supp}},$$

$$tasa_apoyo_autoritarismo = 100 \cdot \frac{\sum s_{democ_support_auth}}{\sum s_{valid_democ_supp}},$$

$$tasa_derecha = 100 \cdot \frac{\sum s_{right_lean}}{\sum s_{valid_lr}}, \quad tasa_izquierda = 100 \cdot \frac{\sum s_{left_lean}}{\sum s_{valid_lr}}.$$

- Para confianza institucional:

$$\text{confianza_congreso_mucha} = 100 \cdot \frac{\sum s_conf_congress_high}{\sum s_valid_conf_congress},$$

$$\text{confianza_congreso_poca_nada} = 100 \cdot \frac{\sum s_conf_congress_low}{\sum s_valid_conf_congress}.$$

- De manera similar se definen tasas de percepción económica, tenencia de bienes y confianza en otras instituciones.
- Las vistas incorporan información nominal de país (`pais_nombre`) y código ISO3 (`pais_iso3`) mediante un mapeo adicional (`country_map`).

1.1.4 Desagregación geográfica y temporal

- **Desagregación por país:** cada fila de `indicadores_latino_pais` corresponde a un par (año, país). Esta vista es el nivel de referencia para análisis comparado entre países.
- **Desagregación por región:** `indicadores_latino_region` añade el código de región dentro de país, permitiendo análisis intra-país (por ejemplo, regiones urbanas vs. rurales, grandes regiones administrativas).
- **Desagregación por ciudad:** `indicadores_latino_ciudad` profundiza hasta el nivel de ciudad/aglomeración, según la definición de `research_city` en cada ola de Latino-barómetro.
- La dimensión temporal está dada por `year`; no hay desagregación mensual, por lo que `periodo=0` en la tabla de sumas.

2 Inventario de indicadores

A partir de las variables normalizadas descritas anteriormente, se construyó un conjunto de indicadores sustantivos para Latinobarómetro. En esta sección se presenta un inventario de los indicadores principales, indicando para cada uno su descripción conceptual y la forma general de cálculo. En todos los casos:

- El denominador se restringe a las personas con **respuesta válida** en la pregunta correspondiente.
- Se utilizan los **factores de expansión** provistos por Latinobarómetro para obtener estimaciones representativas a nivel de país.
- Los resultados se expresan como **porcentajes** (0 a 100).

2.1 Apoyo a la democracia (`democ_supp`)

Descripción. Mide la proporción de personas que declaran que la democracia es *preferible a cualquier otra forma de gobierno*, frente a quienes prefieren un régimen autoritario o manifiestan indiferencia entre ambos tipos de régimen.

Cálculo. A partir de la variable normalizada `democ_supp` se construyen indicadores binarios que distinguen entre:

- Personas que declaran que la democracia es preferible ($s_pref_democ = 1$).
- Personas que prefieren un régimen autoritario.
- Personas indiferentes entre democracia y autoritarismo.

El indicador de apoyo a la democracia se define como el porcentaje de casos con $s_pref_democ = 1$ sobre el total de casos con respuesta válida en esa pregunta, ponderado por el factor de expansión.

2.2 Indiferencia o apoyo a alternativas autoritarias

Descripción. Complementa el indicador anterior cuantificando la proporción de personas que:

1. Declaran que les da lo mismo un gobierno democrático o uno no democrático (indiferencia).
2. Declaran explícitamente que, en algunas circunstancias, un gobierno autoritario puede ser preferible.

Cálculo. Utilizando la misma pregunta normalizada `democ_supp`, se construyen dos proporciones:

- Proporción de personas indiferentes entre democracia y autoritarismo.
- Proporción de personas que prefieren explícitamente un régimen autoritario.

En ambos casos, el denominador son todos los entrevistados con respuesta válida en la pregunta sobre preferencia de régimen.

2.3 Satisfacción con la democracia (`democ_satis`)

Descripción. Mide el grado de satisfacción declarada con el funcionamiento de la democracia en el país, típicamente en una escala de cuatro categorías (por ejemplo, *muy satisfecho*, *más bien satisfecho*, *más bien insatisfecho*, *muy insatisfecho*).

Cálculo. A partir de `democ_satis` se recodifican las categorías en:

- *Alta satisfacción*: respuestas de tipo *muy satisfecho* o *más bien satisfecho*.
- *Baja satisfacción*: respuestas de tipo *más bien insatisfecho* o *muy insatisfecho*.

Se calculan porcentajes de alta y baja satisfacción sobre el total de respuestas válidas, ponderando por el factor de expansión.

2.4 Ubicación ideológica izquierda–centro–derecha (`left_right_scale`)

Descripción. Resume la auto-ubicación ideológica de las personas en una escala izquierda–derecha (usualmente de 1 a 10). Permite estimar la proporción de la población que se identifica como de izquierda, centro o derecha.

Cálculo. Utilizando la variable normalizada `left_right_scale`, se agrupan los valores en tres categorías:

- *Izquierda*: valores bajos de la escala (por ejemplo, 1–4).
- *Centro*: valores intermedios (por ejemplo, 5–6).

- *Derecha*: valores altos (por ejemplo, 7–10).

Para cada grupo se calcula el porcentaje de personas en esa categoría sobre el total de casos con respuesta válida.

2.5 Evaluación de la situación económica del país (`econ_situation`)

Descripción. Captura la percepción sobre la situación económica del país en el momento de la encuesta (*muy buena*, *buena*, *mala*, *muy mala*, u opciones análogas según el año).

Cálculo. A partir de `econ_situation` se recodifican las respuestas en:

- *Visión positiva*: respuestas *muy buena* o *buena*.
- *Visión negativa*: respuestas *mala* o *muy mala*.

Se calculan los porcentajes de visión positiva y negativa sobre el total de respuestas válidas, usando factores de expansión.

2.6 Evaluación de la situación económica personal (`resp_economic_perception`)

Descripción. Mide la percepción que tienen las personas sobre la situación económica de su propio hogar, lo que complementa la evaluación agregada del país.

Cálculo. De manera análoga al indicador anterior, `resp_economic_perception` se recodifica en categorías positivas y negativas. Se calculan los porcentajes de respuesta positiva y negativa sobre el total de casos con respuesta válida, ponderados por el factor de expansión.

2.7 Preocupación por perder el empleo (`job_concern`)

Descripción. Mide el nivel de preocupación de las personas ocupadas respecto a la posibilidad de perder su empleo en el corto plazo.

Cálculo. Se restringe la población a personas que declaran estar ocupadas. A partir de `job_concern` se recodifican las respuestas en:

- *Alta preocupación*: por ejemplo, *muy preocupado*.
- *Baja o nula preocupación*: categorías como *algo preocupado*, *poco preocupado* o *nada preocupado*.

Se calcula el porcentaje de alta preocupación sobre el total de ocupados con respuesta válida.

2.8 Tenencia de bienes y servicios en el hogar (`goods_car`, `goods_hot_water`, `goods_sewage`, `goods_wash_mach`)

Descripción. Conjunto de indicadores que describen el acceso de los hogares a ciertos bienes y servicios básicos, utilizados como aproximaciones a las condiciones materiales de vida:

- `goods_car`: disponibilidad de automóvil en el hogar.
- `goods_hot_water`: acceso a agua caliente.
- `goods_sewage`: conexión a alcantarillado.

- `goods_wash_mach`: disponibilidad de lavadora.

Cálculo. Para cada bien o servicio se construye un indicador dicotómico (1 si el hogar dispone del bien/servicio, 0 en caso contrario). El indicador reporta el porcentaje de hogares que declaran disponer de cada bien, sobre el total de hogares con respuesta válida.

2.9 Confianza en instituciones

Descripción. Agrupa un conjunto de indicadores que miden el grado de confianza en instituciones clave (por ejemplo, Congreso, Gobierno, Poder Judicial, Iglesia, Policía, Fuerzas Armadas, partidos políticos). Aunque los nombres originales de las variables varían por año, todas se normalizan a un esquema común de categorías (*muchas, algo, poca, ninguna confianza*).

Cálculo. Para cada institución se construyen al menos dos indicadores:

- *Confianza alta*: porcentaje de personas que declaran *muchas* o *algo* de confianza en la institución.
- *Confianza baja*: porcentaje que declara *poca* o *ninguna* confianza.

El denominador es el total de personas con respuesta válida para la institución considerada, y los porcentajes se calculan aplicando los factores de expansión de la encuesta.

2.10 Otras variables normalizadas de contexto

Además de los indicadores anteriores, la base armonizada conserva variables normalizadas que se utilizan principalmente como *dimensiones de desagregación* y no como indicadores en sí mismos, tales como:

- `resp_sex`: sexo de la persona entrevistada.
- `resp_age`: edad o grupo de edad.
- `resp_education`: nivel educativo.
- `research_city_size`: tamaño de ciudad o área de residencia.

Estas variables permiten desagregar los indicadores anteriores por género, edad, educación, tamaño de ciudad u otras características sociodemográficas, garantizando comparabilidad temporal gracias al proceso de normalización documentado en el archivo `latinobarometro_glossary_candidata`.

3 Limitaciones y supuestos

- Latinobarómetro es una encuesta de opinión con diseño muestral complejo. La implementación actual no incorpora factores de expansión específicos por diseño; por tanto, los indicadores deben interpretarse como proporciones muestrales ponderadas uniformemente.
- La comparabilidad temporal de las preguntas depende de la estabilidad de la redacción y de la escala de respuesta. Cambios fuertes en las preguntas entre años pueden requerir análisis adicionales.
- La geocodificación a nivel de región/ciudad depende de la calidad y consistencia de los identificadores `research_region` y `research_city`, que pueden variar entre oleadas.