



Trabajo Práctico N°1
Un primer encuentro con la EPH

link del repositorio: <https://github.com/gfeijoo95/BigDataUBA-Grupo7.git>

Grupo 7

Dolimpio, Gastón

Feijoo, Guillermo

Rudi, Federico

Parte I: Familiarizandonos con la base EPH y pobreza

Utilizando información disponible en la página del INDEC, expliquen brevemente cómo se identifica a las personas pobres.

Según la metodología del INDEC, la pobreza se mide a partir de un umbral de ingresos definido por el valor de dos canastas de consumo:

- Canasta Básica Alimentaria (CBA): incluye los alimentos indispensables para cubrir las necesidades nutricionales mínimas de una persona.
- Canasta Básica Total (CBT): amplía la CBA incorporando otros bienes y servicios esenciales, como vestimenta, transporte, educación, salud, entre otros.

A partir de los datos de ingresos que releva la Encuesta Permanente de Hogares (EPH), se compara el ingreso total de los hogares con estos valores de referencia:

- Si un hogar no alcanza a cubrir el valor de la CBT, se considera que sus integrantes están en situación de pobreza.
- Si un hogar no alcanza a cubrir siquiera la CBA, se considera que está en situación de indigencia.

Desde septiembre de 2016, el INDEC publica mensualmente los valores de la CBA y la CBT, lo que permite estimar de manera continua la incidencia de la pobreza y la indigencia en la población urbana del país.

Para el segundo apartado las respuestas se encuentran desarrolladas en el código de la Notebook de Python subida en el repositorio del grupo.

Para el presente ejercicio se toman en cuenta sólo las observaciones del primer trimestre de la región Patagonia para los años 2005 y 2025. De igual forma, el conjunto de variables son las mismas en ambas bases y son las siguientes:

CODUSU: Código para distinguir viviendas

nro_hogar: Código para distinguir hogares

ano4: año

ch04: Sexo

ch06: Edad en años cumplidos

ch07: Situación conyugal

ch08: Tipo de cobertura médica

nivel_ed: Nivel educativo

estado: Condición de actividad

cat_inac: Categoría de inactividad

ipcf: Monto de ingreso per cápita familiar percibido en ese mes

pp04a: ¿El negocio/empresa/institución/actividad en la que trabaja es... (se refiere al que trabaja más horas semanales) 1 = ...estatal? 2 = ... privado? 3 = ...de otro tipo? (especificar)

pondera: Ponderación (factor de expansión)

ID_hogar: Identificador a nivel de hogar

iv1: Tipo de vivienda

iv2: ¿Cuántos ambientes/habitaciones tiene la vivienda en total?

iv6: Tiene agua...

iv8: ¿Tiene baño/letrina?

IX_Tot: Cantidad de miembros del Hogar

(Opcional) De la base de datos de Hogares, de ambos años, seleccione y limpie a un único formato las siguientes variables: IV1, IV2, IV6, IV8 e I18. Pruebe, unir la base de datos Individuales con Hogares usando CODUSU y NRO_HOGAR, usando los cuatro tipo de uniones que vimos en la clase tutorial. Comente y reporte la siguiente tabla:

La tabla 1 reporta el resumen de la estructura de la base resultante de la unión entre la base a nivel de individuo y de hogar. Para ello, se llevó a cabo un proceso de limpieza de datos que incluyó la eliminación de espacios en blanco y caracteres especiales (para CODUSU y NUM_HOGAR) en ambas bases y para ambos años.

El número de filas y columnas no varía, dado que el mismo procedimiento se aplicó tanto a la base de hogares como a la de individuos. Es decir, antes de la unión se eliminaron las observaciones correspondientes a regiones distintas de la Patagonia y se conservaron únicamente las variables requeridas para el ejercicio, junto con la variable que indica el tamaño del hogar.

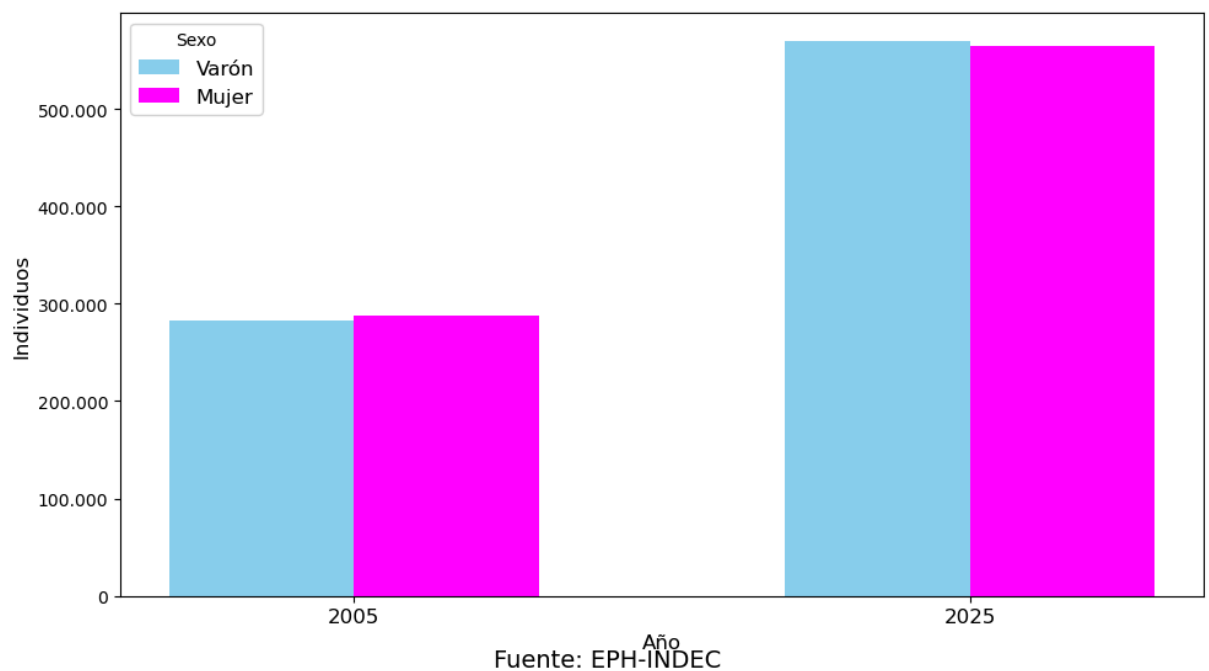
Tabla 1. Reporte de unión de base Hogares e Individuos de la EPH

Tipo de unión	Nro. de filas		Nro. de Columnas		Total de NAs	
	1T 2005	1T 2025	1T 2005	1T 2025	1T 2005	1T 2025
unión por intersección	3172	5326	20	20	0	0
unir por izquierda	3172	5326	20	20	0	0
unir por derecha	3172	5326	20	20	0	0
unión conjunta	3172	5326	20	20	0	0

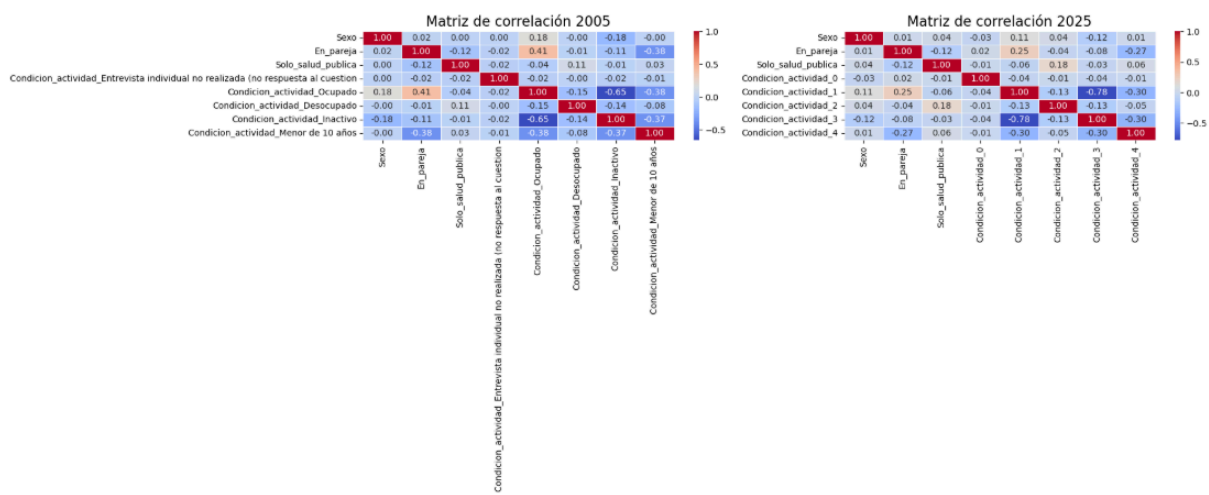
Parte II: Análisis Exploratorio

En la comparación entre los veinte años transcurridos entre 2005 y 2025 se observa en primer lugar que la población se duplicó en la región y que la de varones aumentó ligeramente más que la de las mujeres, pasando de ser minoritarios a mayoritarios en el período señalado.

Población por sexo en la Patagonia



Fuente: EPH-INDEC



Parte III: Conociendo a los pobres y no pobres

1. Identificación de hogares con información faltante

Uno de los principales desafíos de la Encuesta Permanente de Hogares (EPH) es la falta de respuesta en variables clave, particularmente en el ingreso total familiar (ITF) y en la condición de actividad.

En el caso de la región **Patagónica**, el análisis revela que alrededor del **44% de las personas no reportaron su condición de actividad**, mientras que aproximadamente **1.225 observaciones (14%)** no informaron su ingreso total familiar.

Este nivel de no respuesta constituye una limitación importante para el análisis de pobreza, ya que puede sesgar la estimación de ingresos y subestimar la magnitud del fenómeno.

Para continuar con el trabajo, se conformaron dos subconjuntos:

- **respondieron**: incluye las observaciones con un ITF reportado y positivo.
- **norespondieron**: agrupa aquellas con ITF igual a cero.

Este filtrado garantiza que el análisis posterior se realice únicamente sobre los hogares con información económica válida.

2. Cálculo de adultos equivalentes

A partir de la tabla de **adultos equivalentes**, se asignó un peso relativo a cada persona según su **edad y sexo**, reflejando las diferentes necesidades de consumo energético.

Por ejemplo, un varón de 2 años equivale a 0,46 adultos, mientras que una mujer adulta representa 1 adulto equivalente.

Luego, se sumaron los valores de cada individuo dentro del hogar para construir la variable **AD_EQUIV_HOGAR**, que mide la cantidad total de adultos equivalentes por hogar.

En promedio, los hogares patagónicos presentan entre **2,5 y 3,3 adultos equivalentes**, aunque se observan casos extremos de hasta más de 8 adultos equivalentes, lo que sugiere una heterogeneidad significativa en la composición familiar de la región.

3. Ingreso necesario y línea de pobreza

El siguiente paso consistió en calcular el **ingreso necesario** para no ser pobre, multiplicando la **Canasta Básica Total (CBT)** por el número de adultos equivalentes del hogar.

Se consideraron los siguientes valores de referencia:

- **CBT 1er trimestre 2005**: \$205,07

- **CBT 1er trimestre 2025: \$365.177**

De este modo, el **ingreso mínimo requerido** varía según el tamaño y composición del hogar: un hogar de 3 adultos equivalentes necesitaba en 2005 al menos **\$615** para superar la línea de pobreza, mientras que en 2025 el valor asciende a más de **\$1.095.000**. Estos valores reflejan el fuerte incremento del costo de vida en el período analizado.

4. Clasificación de pobreza

Con las variables anteriores se construyó la variable **pobre**, que toma valor **1** si el ingreso total familiar (ITF) es menor al ingreso necesario, y **0** en caso contrario.

El análisis muestra que en **2005**, el **18,4% de los hogares** patagónicos eran pobres, mientras que en **2025** la proporción ascendió a **21,1%**.

Esto indica un **aumento de casi 3 puntos porcentuales** en la tasa de pobreza en los últimos veinte años, lo que sugiere un deterioro relativo en el poder adquisitivo de los hogares frente al costo de vida.

5. Estadísticas descriptivas

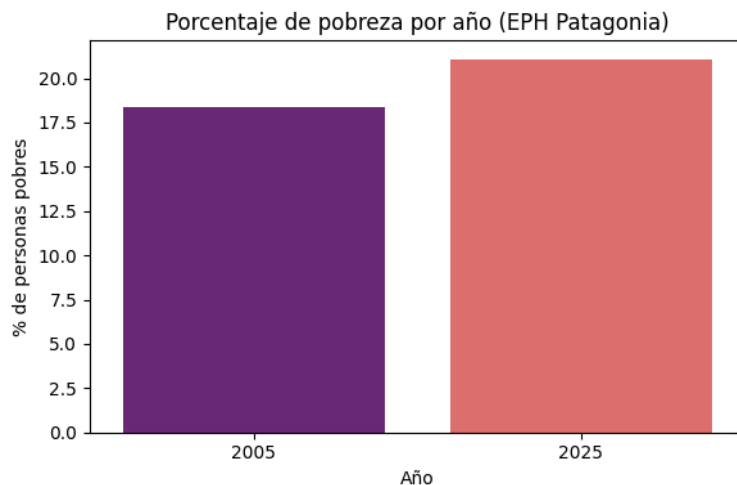
Al desagregar la información por condición de pobreza, se observa que los **hogares pobres** presentan ingresos familiares significativamente inferiores y tamaños promedio de hogar mayores, medidos en adultos equivalentes.

En contraste, los **hogares no pobres** concentran los niveles más altos de ingreso, con menor dispersión y estructura demográfica más acotada.

Esto refleja la persistencia de una brecha estructural de ingresos dentro de la región, incluso entre hogares con composiciones familiares similares.

6. Análisis gráfico

Gráfico 1 – Porcentaje de pobreza por año (EPH Patagonia)

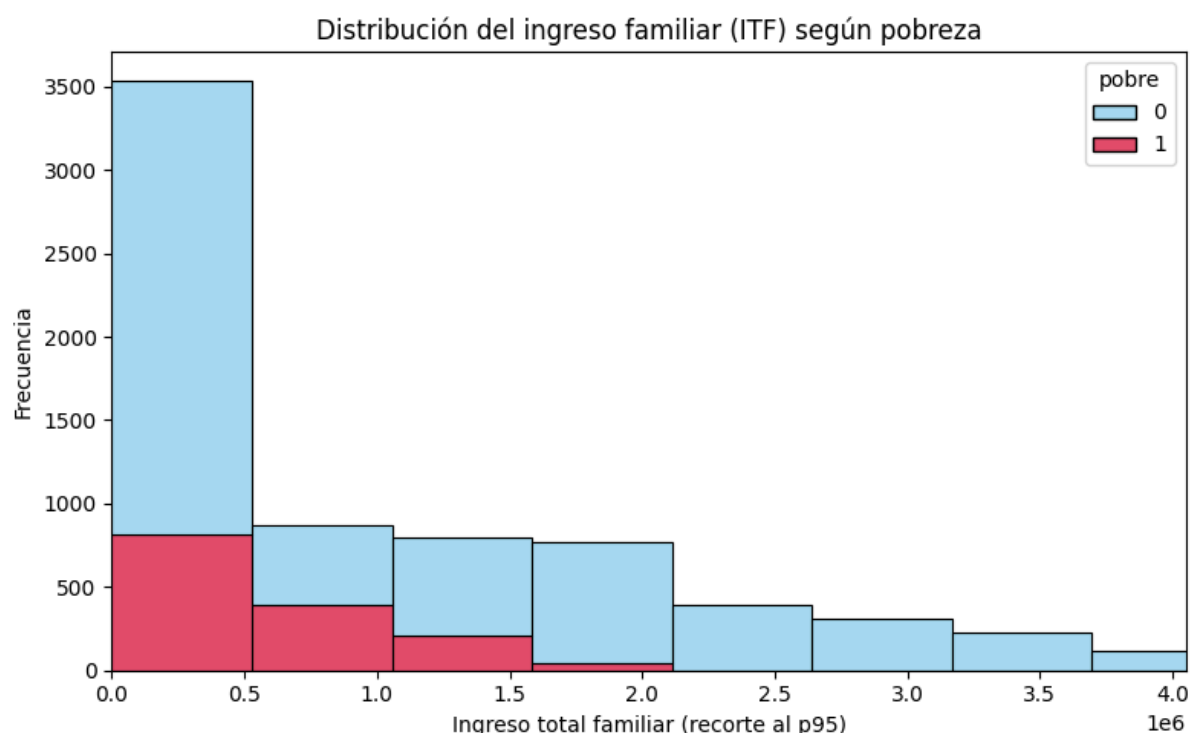


El primer gráfico muestra la evolución del porcentaje de pobreza entre 2005 y 2025. Se aprecia un **incremento leve pero sostenido** en la proporción de hogares bajo la línea

de pobreza, que pasa de cerca del **18% en 2005** a **21% en 2025**.

Este aumento, aunque moderado, resulta significativo en términos de bienestar social, ya que se da en un contexto de crecimiento general del ingreso nominal pero también de una mayor inflación y aumento de precios básicos.

Gráfico 2 – Distribución del ingreso familiar (ITF) según pobreza



El segundo gráfico representa la distribución del **ingreso total familiar (ITF)**, recortado al percentil 95 para evitar valores extremos.

La concentración de observaciones en los tramos inferiores evidencia que los **hogares pobres** (en color rojo) se concentran mayoritariamente por debajo del primer segmento de ingresos, mientras que los **no pobres** (en celeste) presentan una distribución más dispersa hacia valores altos.

Esta diferencia visual confirma la fuerte asimetría en la distribución del ingreso: la mayor parte de la población tiene ingresos cercanos al umbral de pobreza, y solo una fracción minoritaria logra superar holgadamente ese límite.

7. Conclusiones generales de la Parte III

El análisis de la EPH para la región Patagónica muestra que, entre 2005 y 2025, **la pobreza no se redujo de manera significativa**, a pesar del incremento de los ingresos nominales.

La línea de pobreza se elevó de forma pronunciada por el aumento del costo de la canasta básica, y una porción importante de hogares permanece vulnerable ante variaciones en los precios o ingresos laborales.

En síntesis, el estudio refleja la **persistencia de la pobreza estructural** en la región, así como la necesidad de políticas redistributivas que acompañen el crecimiento de los precios y mejoren la capacidad de compra real de los hogares.