



UNIVERSIDAD DE  
BUENOS AIRES – FCE

## ***TALLER DE PROGRAMACIÓN***

### **TP N°1 UN PRIMER ENCUENTRO CON LA EPH**

**PROFESORA: María Noelia Romero**

**Alumnos: Bergant, Agustín Rivas  
Cabrera, Juan  
Leiva, Francisco**

## Introducción

El presente trabajo busca dar respuesta a las consignas del primer trabajo práctico del seminario “Taller de programación”, a cargo de María Noelia Romero, perteneciente a la Maestría en economía aplicada. Entre los propósitos buscados se incluyen: introducir a los conceptos básicos de elaboración de tablas y gráficos en Python, así como la manipulación de bases de datos, empleando para ello la Encuesta Permanente de Hogares (EPH), confeccionada por el INDEC. De lo anterior se desprende que todas las figuras y tablas desarrolladas en el presente trabajo proviene de dicha fuente, eliminando en consecuencia la referencia en cada una de ellas.

1) El Instituto Nacional de Estadística y Censos (INDEC) identifica a las personas pobres mediante un método indirecto basado en los ingresos de los hogares. En lugar de medir la pobreza a través de condiciones de vida o acceso a bienes, el organismo compara los ingresos declarados por las familias con el costo de dos canastas de consumo: la Canasta Básica Alimentaria (CBA) y la Canasta Básica Total (CBT). La primera refleja el valor monetario necesario para cubrir un conjunto mínimo de alimentos que satisfacen requerimientos energéticos y proteicos esenciales, mientras que la segunda amplía esa cobertura e incorpora otros bienes y servicios indispensables (como indumentaria, transporte, salud, educación y vivienda).

Un hogar se considera indigente cuando sus ingresos totales no alcanzan para cubrir la CBA. En cambio, se clasifica como pobre pero no indigente si sus ingresos superan el valor de la CBA pero son insuficientes para cubrir la CBT. En términos prácticos, la línea de indigencia marca el umbral mínimo para la subsistencia alimentaria, mientras que la línea de pobreza delimita el ingreso necesario para satisfacer un nivel básico de vida.

Para realizar estas comparaciones, el INDEC ajusta los ingresos de cada hogar según su composición mediante la noción de “adultos equivalentes”, que pondera las necesidades relativas de acuerdo con la edad, el sexo y la cantidad de integrantes. Los datos provienen de la Encuesta Permanente de Hogares (EPH), que releva periódicamente información sobre ingresos, empleo y características sociodemográficas en las principales áreas urbanas del país. Con esta información, el organismo estima los porcentajes de hogares y personas cuyos ingresos se sitúan por debajo de las líneas de pobreza e indigencia, permitiendo así monitorear la evolución de la pobreza monetaria en la Argentina.

## Parte I

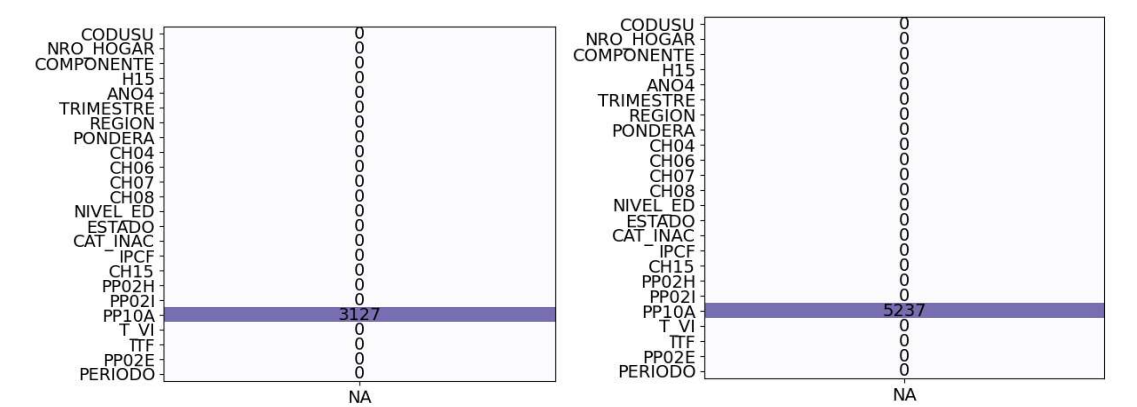
A continuación, se detalla el proceso llevado a cabo en la depuración de la base de datos.

En primer lugar, se seleccionó la Región Patagónica como unidad de análisis, por lo que se eliminaron los registros correspondientes a las demás regiones del país. Posteriormente, se unieron los datos de los trimestres 2005 y 2025 en una única base consolidada para realizar el análisis comparativo. Las 15 variables seleccionadas, para la encuesta a individuos, son: sexo del encuestado/a (CH04), años cumplidos (CH06), estado civil (CH07), tipo de cobertura medica (CH08), nivel educativo (NIVEL\_ED), condición de actividad laboral (ESTADO), categoría o motivo de inactividad (CAT\_INAC), monto del ingreso per capita familiar (IPCF), lugar de nacimiento (CH15), motivo por el que no estuvo buscando trabajo los últimos 30 días (PP02E), si buscó o no trabajo los últimos 12 meses (PP02H), si trabajó o no en los últimos 12 meses (PP02I), hace cuánto tiempo está buscando trabajo (PP10A), monto total de ingresos no laborales (T\_VI) y monto del ingreso total familiar (ITF).

Se procedió a recodificar las variables siguiendo las categorías de la base del año 2025, con el fin de garantizar la comparabilidad entre ambos períodos. Asimismo, se realizó una limpieza de datos, corrigiendo o eliminando aquellas observaciones que generaban ruido o resultaban inconsistentes, como los casos de edades negativas y valores del IPCF en negativo, los cuales no tienen sentido económico. Finalmente, se verificó que ambas bases compartieran los mismos tipos de variables y formatos para cada campo, asegurando la correcta integración y posterior análisis. La Figura 1 presenta un heatmap con la cantidad de valores faltantes por cada una de las 15 variables para los dos años en cuestión. A simple vista, puede observarse que la única variable con observaciones nulas

es PP10A, que es la cantidad de tiempo desde que la persona encuestada está buscando trabajo. Para el resto de variables, la muestra está completa en ambos trimestres.

Figura 1: Cantidad de valores faltantes por variable.



Por su parte, para la encuesta a hogares, las variables seleccionadas fueron: tipo de vivienda (IV1), cantidad de ambientes/habitaciones de la vivienda (IV2), el tipo de acceso al agua que tiene la vivienda (IV6), si la misma posee baño o letrina (IV8), y el tipo de energía (gas en red, gas en tubo/garrafa o kerosene/leña/carbón) utilizada para cocinar (II8). De manera similar al tratamiento aplicado en la base de individuos, se recodificaron los valores de ciertas variables con el propósito de asegurar la comparabilidad entre los dos años analizados, para luego unificar ambas bases. El objetivo final fue vincular esta base con la de individuos mediante las variables clave CODUSU y NRO\_HOGAR, permitiendo integrar la información a nivel de hogar y persona.

Por último, en esta primera parte del trabajo, se realizó la unión de las bases de individuos y hogares. Para ello, se utilizaron los 4 tipos disponibles de merge o unión: por intersección, por izquierda, por derecha o conjunto. El objetivo es revisar el funcionamiento de esta herramienta y analizar la estructura de las bases resultantes. A continuación, se presentan los valores finales de filas, columnas y datos faltantes por tipo de unión.

Cuadro 1. Resultados por tipo de unión de las bases individuos y hogares.

| Tipo de unión          | Filas | Columnas | Faltantes |
|------------------------|-------|----------|-----------|
| Unión por intersección | 8551  | 35       | 8342      |
| Unión por izquierda    | 8588  | 35       | 8748      |
| Unión por derecha      | 8551  | 35       | 8342      |
| Unión conjunta         | 8588  | 35       | 8748      |

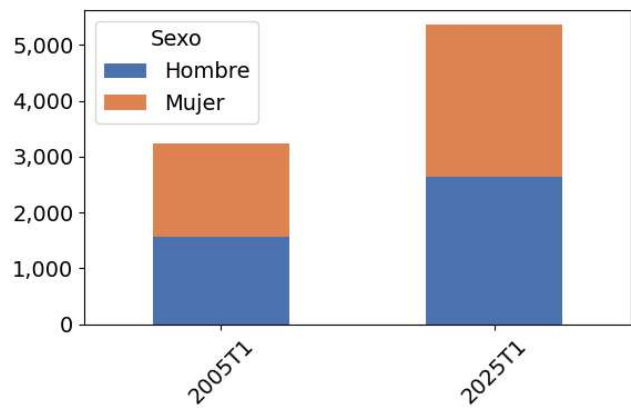
Debido a que la base de los hogares tiene menos registros totales, hay menos hogares que individuos. La unión por intersección y la unión por derecha coinciden en cantidad de filas, columnas y datos faltantes. La base de los hogares, en estos dos casos, hace de “cuello de botella” a la hora de integrar una base con otra (*merging process*). Para la unión por derecha, los datos de los hogares son los utilizados como base . Por último, y por razones similares a la recién descrita, la unión por conjunto y por izquierda también contienen la misma información de filas, columnas y datos faltantes. En este tipo de uniones, la base de individuos, con mayor cantidad de filas, es la utilizada como referencia. Los datos faltantes, para todos los tipos de unión están, en su gran mayoría, explicados por una sola variable.

### Parte II

A continuación, se procede a comenzar con el análisis exploratorio de los datos de estudio. Como primer paso de esta inspección de los datos se hace foco en la composición por sexo de la Región

Patagonia y como varió entre la encuesta del primer trimestre del 2005 y la del primer trimestre del 2025.

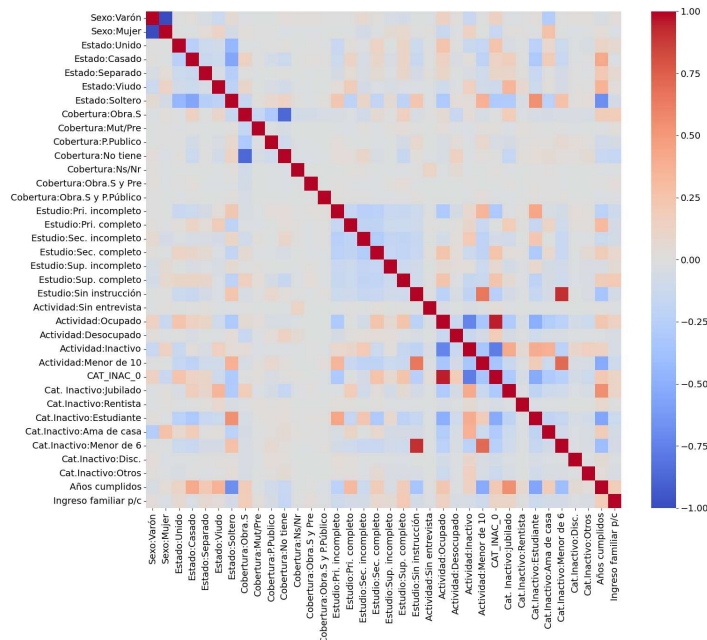
Figura 2 | Cantidad de personas por sexo según período



Como puede observarse a simple vista, la primera diferencia importante se manifiesta entre la cantidad de personas relevadas en uno y otro año. A las 3329 personas encuestadas en el 2005, se le suman 2130 más para el mismo trimestre del año 2025. Es decir, como primer hecho descriptivo puede comentarse que aumentó el tamaño muestral entre los periodos analizados. En cuanto a la composición por sexo, las proporciones son prácticamente similares. En el 2005, el 48,7% de las personas encuestadas eran varones, mientras que en el 2025 la proporción aumentó menos de 1 punto porcentual (p.p) y llegó a 49,4% del total.

Con el fin de seguir con la investigación exploratoria, se analizarán las correlaciones resultantes de las siguientes variables: CH04 (sexo), CH06 (años cumplidos de la persona), CH07 (estado civil de la persona), CH08 (tipo de cobertura médica), NIVEL\_ED (nivel educativo máximo alcanzado), I18 (estado de actividad de la persona), CAT\_INAC (categoría de inactividad) y IPCF (ingreso familiar per capita). Para poder generar la matriz, se realizaron modificaciones en las variables categóricas empleadas, generando variables dicotómicas o dummies para representar cada categoría de cada variable.

Figura 3 | Matriz de correlación entre las variables seleccionadas



En líneas generales, la mayoría de las correlaciones resultantes son débiles, con un nivel de correlación dentro del rango 0,3 a -0,3, existiendo resultados de mayor magnitud entre categorías que forman parte de la misma variable. Sin embargo, hay algunos aspectos interesantes a destacar. De forma intuitiva, el mayor nivel educativo posible, superior completo, correlaciona positivamente con estar empleado. A su vez, las personas sin instrucción tienden, aunque de forma débil, a tener menores ingresos familiares per cápita. Lo que sucede con la cobertura médica también es interesante aunque las magnitudes sean pequeñas: las personas con obra social y prepagas/mutuales correlacionan positivamente con ingreso familiar per cápita, mientras que las que poseen planes públicos o no tienen cobertura lo hacen de forma negativa. Por último, y como posible síntoma de la brecha laboral de género, las mujeres correlacionan positivamente con la actividad ama de casa.

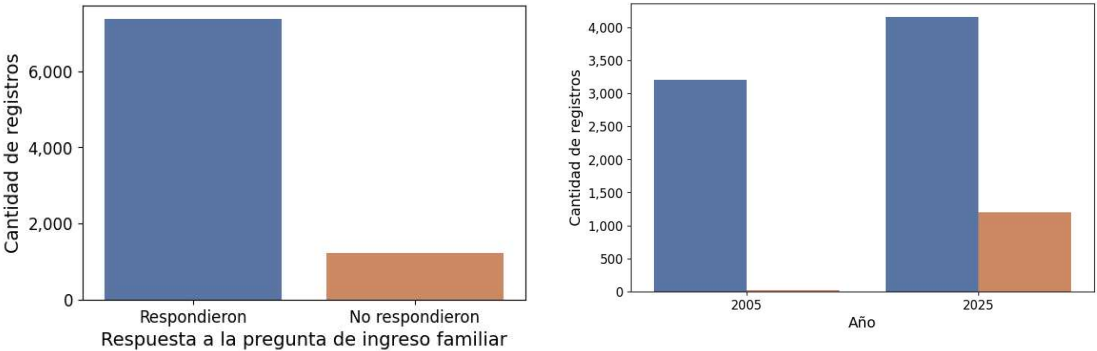
**Parte III**

Dentro de esta sección se realizan comparaciones entre las bases de datos de 2005 y la de 2025, pudiendo encontrar de esta manera diferencias como producto de una comparación entre dos décadas.

En primer lugar, interesó explorar la cantidad de registros que responden la pregunta acerca de la actividad principal. En rigor de verdad, la pregunta indica la cantidad de dinero que ingresa en la familia, aproximadamente. Sobre un total de 8.588 registros, el 85,7% (7.363 registros) respondieron un ingreso determinado del hogar, lo que en los términos actuales significa haber respondido la pregunta acerca de la actividad económica. En contraposición, 1.225 registros omitieron responder a esta pregunta.

No obstante, este primer acercamiento, que puede verse del lado izquierdo de la figura X, lo cierto es que es significativamente diferente el grado de respuesta según se ponga la atención al año 2005 o bien el año 2025. Para ello, podrá ayudarse de la figura del lado derecho: esta vez teniendo en cuenta la negativa: sólo 23 de 3.229 encuestados (0,7%) no respondieron la pregunta del ingreso total familiar en 2005 *versus* 1.202 sobre 5.359 (22,4%) que no respondieron en 2025. En términos de hogares la diferencia es similar: el 1,2% de los hogares de la región patagónica no respondieron la pregunta (11 sobre 949 hogares) mientras que el 20,3% (422 de 2.083 hogares) no respondieron en 2025.

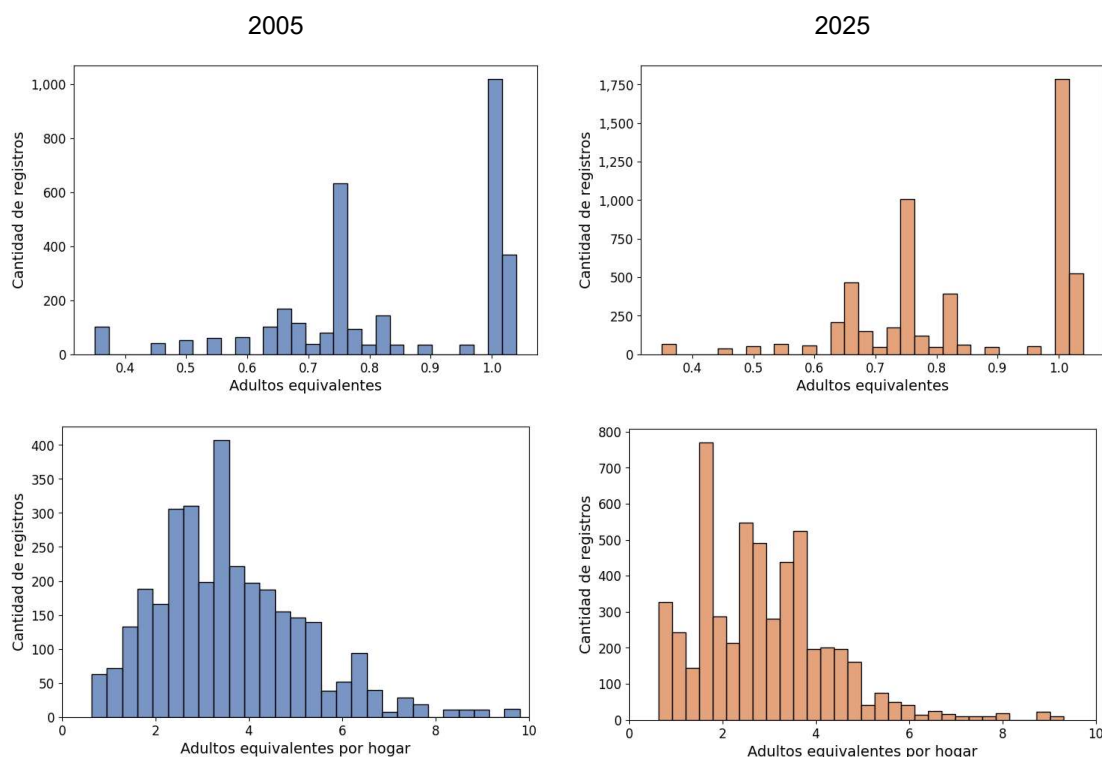
**Figura 4 | Cantidad de registros que respondieron la pregunta. Año 2005 vs 2025.**



Para finalizar con el análisis, se guardaron los registros en dos objetos, denominados 'respondieron' y 'norespondieron', según si contestaron o no la pregunta relacionada con la actividad económica familiar. A continuación, se estudió la cantidad de adultos equivalentes por hogar, utilizando el archivo "tabla\_adulto\_equiv.xlsx", el cual permite transformar cada registro individual en un adulto, considerando la edad y las necesidades energéticas correspondientes a un adulto de dicha edad.

Los resultados, para uno y otro año, se observan en la Figura X, destacando una distribución bastante similar a nivel registro (véase paneles superiores), aunque un poco diferente cuando se considera a nivel hogar (véase paneles inferiores)

**Figura 5 | Distribución de adultos equivalentes y adultos equivalentes totales en el hogar.**



A continuación, se incorporó un umbral de referencia directo para la clasificación de hogares en condición de pobres y no pobres. Para ello, se empleó el valor de la CBT de cada período: aproximadamente \$205,07 y \$365.177 para 2005 y 2025, respectivamente. A partir de estos valores, se creó en la base de registros individuales una nueva variable, denominada 'ingreso\_necesario', que ajusta la CBT a las características de cada hogar. Posteriormente, esta variable permitió construir la variable dicotómica (o *dummy*) 'pobre', la cual toma el valor 1 si el ingreso familiar total del hogar es INFERIOR a este umbral (ingreso\_necesario) y 0 en caso contrario.

Los datos indican que la incidencia de la pobreza ha experimentado un aumento entre los períodos analizados. En rigor, en el año 2005, el 18,7% de los hogares se encontraban en condición de pobreza, mientras que en 2025, este porcentaje se incrementó a 22,5%, constituyendo un aumento de 3,8 puntos porcentuales en la tasa de pobreza en dos décadas. A su turno, este incremento sugiere un deterioro en la capacidad de los ingresos familiares para cubrir la CBT, lo que implica que una mayor proporción de la población ha caído o se ha mantenido por debajo del umbral de la Línea de Pobreza. A continuación, se presenta la Tabla X como un resumen de las estadísticas descriptivas.

**Cuadro 2 | Estadísticas descriptivas de registros pobres.**

| Año  | Registro pobres | Registros totales | Porcentaje |
|------|-----------------|-------------------|------------|
| 2005 | 600             | 3.206             | 18,7%      |
| 2025 | 934             | 4.157             | 22,4%      |

Se incorporaron dos indicadores adicionales para analizar la incidencia de la pobreza en la Patagonia de manera desagregada: por nivel educativo y por tamaño del hogar.

El primer indicador, detallado en la Figura 6, examina la distribución de la pobreza por Nivel Educativo y revela una vulnerabilidad diferencial. Se observa un crecimiento generalizado en la tasa de pobreza en 2025 en casi todas las categorías educativas en comparación con 2005. Este incremento es notable en los grupos de Secundaria Incompleta, cuya tasa asciende de un  $\approx 22\%$  a un  $\approx 32\%$ , y en los niveles Secundaria Completa y Superior Universitaria Incompleta, donde las tasas casi se duplican. La única excepción es Primaria Completa, donde la tasa de 2025 ( $\approx 20\%$ ) es ligeramente inferior a la de 2005 ( $\approx 22\%$ ), aunque se mantiene el patrón de que la menor formación conlleva la mayor incidencia de pobreza.

En segundo lugar, el análisis del cruce entre el tamaño del hogar y la proporción de pobres se presenta en la Figura 7, la cual evidencia una relación fuertemente positiva entre la cantidad de miembros del hogar y el riesgo de pobreza. La línea de 2025 muestra consistentemente una tasa de pobreza superior a la de 2005 para todos los tamaños de hogar. La vulnerabilidad es extrema en las familias numerosas: para hogares con 11 o más miembros, la tasa de pobreza alcanza el 100% en ambos periodos, lo que subraya la incapacidad de los ingresos familiares para cubrir la CBT ajustada.

Figura 6 | Porcentaje de registros pobres en función del nivel educativo

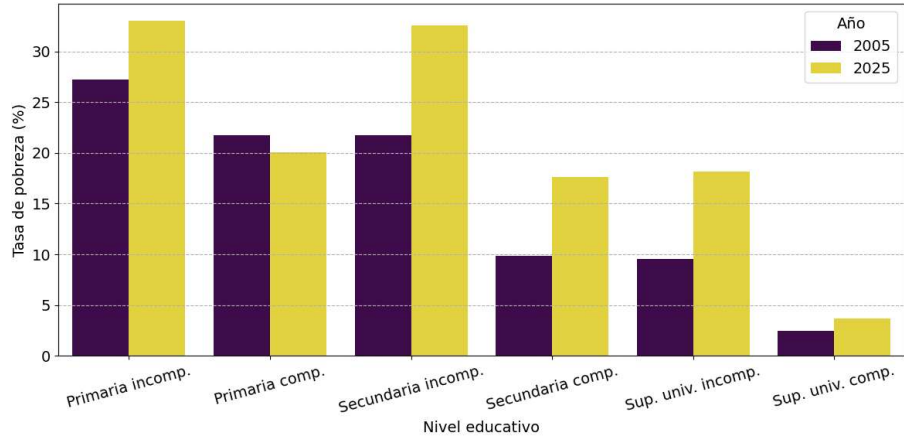


Figura 7 | Porcentaje de hogares pobres en función del tamaño del hogar

