



Universidad de Buenos Aires  
Facultad de Ciencias Económicas

Universidad de Buenos Aires.  
Facultad de Ciencias Económicas  
Escuela de Estudios de Posgrado.

**MAESTRÍA EN ECONOMÍA APLICADA**

Taller de Programación

**Primer Trabajo Práctico:**

**Un primer encuentro con la EPH**

**AUTORES:**

KOSTZER FEDERICO  
MEZA PALMA MANUEL H.  
PARDINI MAXIMILIANO G.

**DOCENTE:** ROMERO MARIA NOELIA

OCTUBRE 2025

## Parte I: Familiarizándonos con la base EPH y limpieza

En Argentina el INDEC viene realizando estimaciones sistemáticas y periódicas de la incidencia y otras características de la pobreza desde principio de los años 90 en el Gran Buenos Aires, la preparación de estos indicadores se ha extendido a partir de 2001 mediante ciertas ampliaciones metodológicas para el conjunto de las áreas urbanas representadas por la cobertura de la Encuesta Permanente de Hogares.

Según consigna la institución en sus documentos metodológicos la definición de pobreza está dada de la siguiente forma:

El INDEC utiliza para esas estimaciones el “enfoque del ingreso” siguiendo el concepto de pobreza absoluta. De acuerdo con este método, un hogar es considerado pobre si su ingreso (esto es, la suma de los ingresos que reciben todos los miembros del hogar) resulta menor que la “línea de pobreza”. Este último es un concepto normativo pues representa el valor de todos los bienes y servicios que se consideran necesarios para que el hogar satisfaga las necesidades básicas. (INDEC, s.f., p. 2)

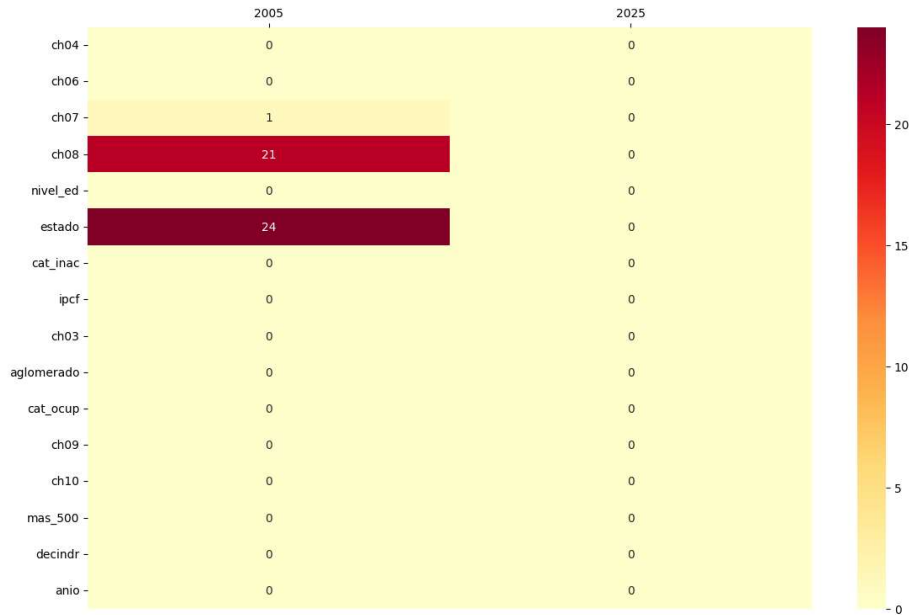
La región elegida para realizar el estudio fue la Pampeana que registró 14.651 observaciones para 2005 y 13.803 en el primer trimestre de 2025 quedando la base total con 28.454 observaciones y 245 variables de estas para el resto del trabajo se eligieron 15 que abarcan características sociodemográficas y laborales de los individuos las cuales se detallan en la tabla 1.

**Tabla 1. Variables elegidas**

Variable	Descripción	Tipo de dato
ch04	Sexo	num
ch06	Edad	num
ch07	Estado civil	cat
ch08	Cobertura médica	cat
nivel_educ	Nivel educativo	cat
estado	Condicion de empleo del entrevistado	cat
cat_inac cat	Categoría de inactividad	cat
ipcf num	Monto del ingreso per cápita familiar	num
ch03	Relación de parentesco	cat
aglomerado	Aglomerado	cat
cat_ocup	Categoría ocupacional	cat
ch09	Alfabetismo	cat
ch10 cat	Asistencia a establecimiento educativo	cat
mas_500 cat	Aglomerados según tamaño poblacional	cat
decindr cat	Nº de decil de ingreso total individual del total EPH	cat

En el proceso de limpieza de esta base final se comprobó que para el año 2005 existían muchas variables que presentaban alta heterogeneidad en cuanto al tipo de dato codificado lo que generaba un volumen alto de valores faltantes; luego de operaciones para estandarizar los tipos de datos se redujo de manera considerable esta situación obteniéndose los resultados de la figura 1.

Figura 1. Cantidad de valores faltantes por variable y año

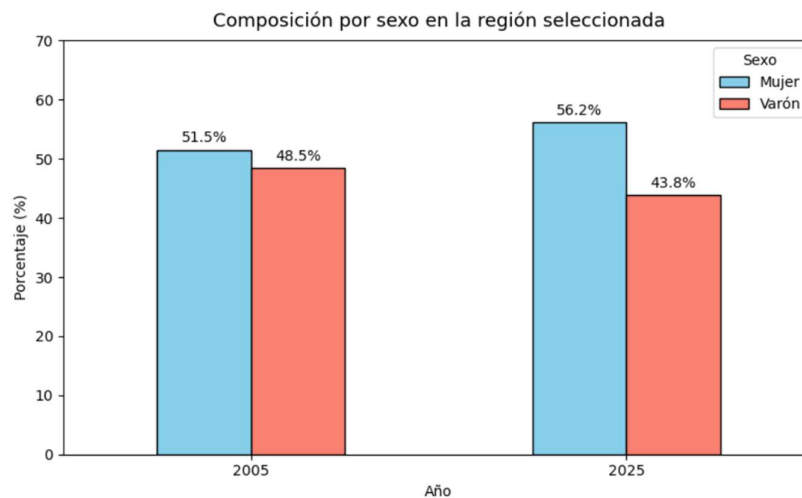


Para completar la limpieza de la base se realizó una verificación de valores extraños sobre todo en las variables numéricas como ingreso que pueden presentar cifras que no tienen sentido como el caso de números negativos y otro tipo de codificación que suele indicar condiciones de no respuesta o no aplicabilidad; la ejecución arrojó efectivamente valores extraños que se computaron como faltantes; la variable más afectada fue la edad (ch06) que acumuló 297 casos seguida por estado civil (ch07) con un caso, lo demás se mostró inalterado.

## Parte II: Primer Análisis Exploratorio

### 3. Composición por sexo (2005 vs. 2025)

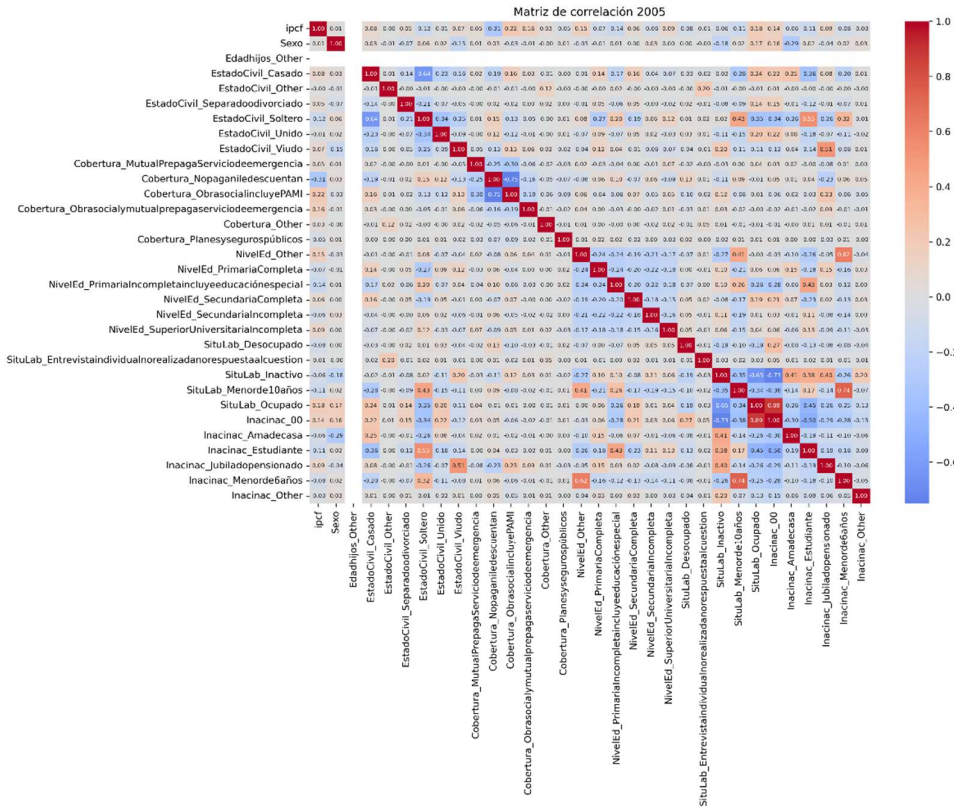
En la región pampeana, se observa un incremento en la proporción de mujeres dentro de la muestra, pasando del 51,5% en 2005 al 56,2% en 2025, lo que implica un aumento de 4,7 puntos porcentuales. En contraposición, la participación de varones disminuyó del 48,5% al 43,8% en el mismo período.



Estos resultados sugieren una mayor presencia relativa de mujeres en la estructura poblacional reciente, posiblemente vinculada a diferencias en la composición etaria, la participación laboral o la cobertura de la encuesta en la región.

#### 4. Matriz de correlaciones sociodemográficas y laborales: 2005 vs 2025

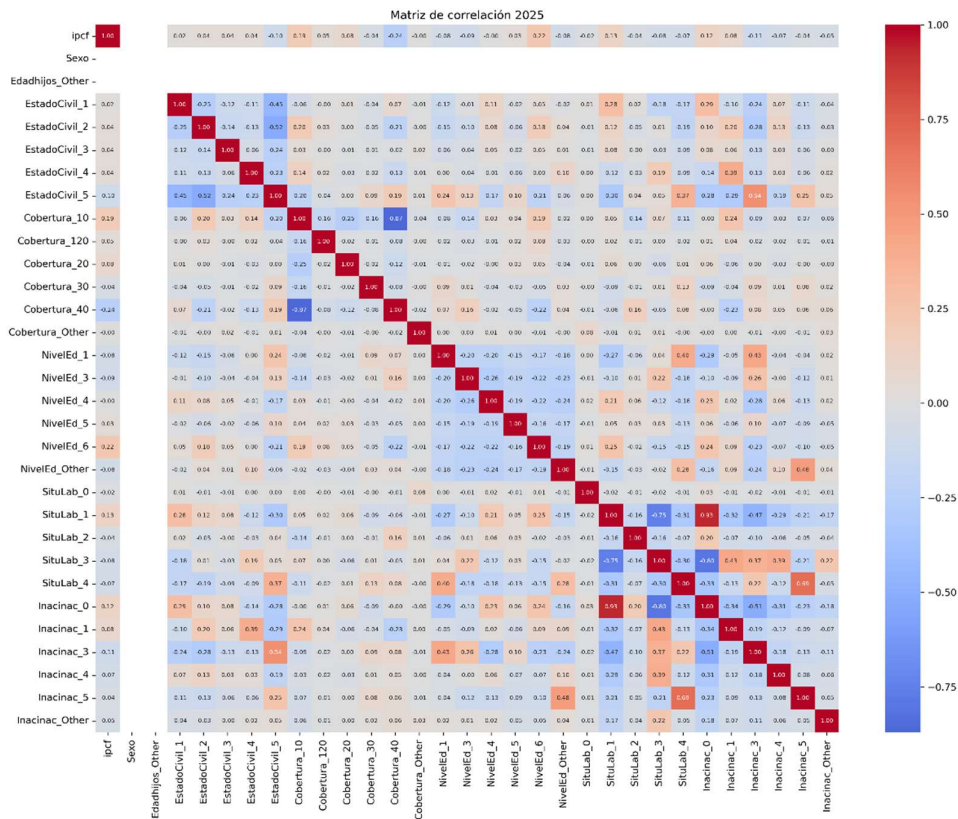
Las correlaciones más altas se observan entre las variables que capturan la situación laboral (SituLab) y la condición de inactividad (Inacinac). En 2005, la fuerte asociación entre *SituLab\_Ocupado* e *Inacinac\_00* ( $r=0.89$ ) refleja la lógica interna del mercado de trabajo: quienes se declaran ocupados tienden a no pertenecer a ninguna categoría de inactividad, es decir, existe una coherencia estructural entre ambas codificaciones. En 2025, esta relación se mantiene e incluso se refuerza (*SituLab\_1*–*Inacinac\_0*,  $r=0.93$ ), lo que sugiere una mayor consistencia en la clasificación laboral y una menor ambigüedad en las respuestas del relevamiento, posiblemente por mejoras en la captación o en la claridad de las categorías ocupacionales.



Por otra parte, las altas correlaciones entre variables de cobertura sanitaria ( $\approx 0.75$  en 2005 y  $\approx 0.87$  en 2025) indican que las personas que cuentan con un tipo de cobertura —por ejemplo, obra social o descuentos en farmacia— tienden a acumular otras formas de protección. Este patrón sugiere una segmentación positiva del bienestar, donde los individuos integrados al empleo formal o con mayores recursos presentan múltiples coberturas simultáneas.

En conjunto, los resultados muestran una estructura más definida y polarizada entre los grupos ocupados/inactivos y una mayor consolidación de la cobertura sanitaria

entre quienes ya acceden al sistema, reflejando tanto mejoras institucionales en la medición como persistencia de brechas estructurales en el acceso a empleo y salud.



### Parte III: Conociendo a los pobres y no pobres

#### No respuesta de ingresos en la encuesta permanente de hogares

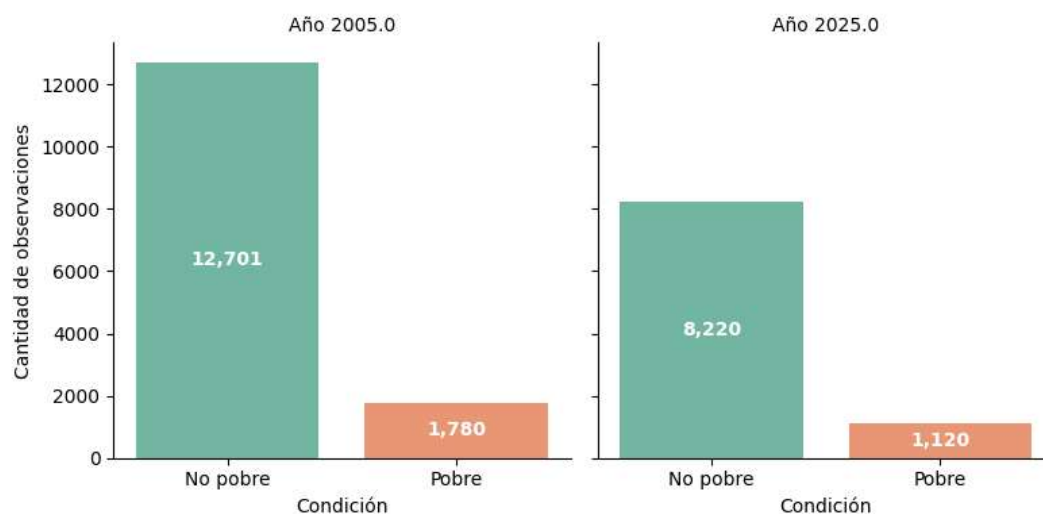
Como señala el documento técnico del Ministerio de Hacienda en el siguiente [informe](#), la proporción de respuestas respecto al ingreso total familiar fue disminuyendo a lo largo de los años.

Si comparamos el año 2005 con 2025 la cantidad de personas que no respondieron esta pregunta es 28 veces mayor. En el año base para la región de análisis, de los 14.651 hogares solamente 170 decidieron no responder, dándonos un 98% de registros no nulos; mientras que en 2025 de los 13.803 hogares hubo un total de 4.463 que decidieron no responder, dándonos un 67% de hogares que respondieron y casi un tercio que decidió no hacerlo.

#### Conociendo a los pobres y no pobres

Para medir la pobreza en este apartado lo que se realizó fue comparar los ingresos de una familia con la canasta básica total en el año del registro multiplicado por la cantidad de adultos equivalente en el hogar y el estado de pobre o no pobre será la misma para todos los miembros del hogar.

Ahora compararemos los resultados obtenidos de nuestra base de datos sobre las personas pobres y no pobres en la región pampeana en el año 2005 y 2025.



Como vemos en la figura anterior, para el año 2005 tenemos 12.701 no pobres y 1.780 pobres, dándonos un porcentaje de 87,7 no pobres y 12,3 pobres. Mientras que para 2025 tenemos 8.220 no pobres y 1.120 pobres lo que resulta en un 88% de no pobres y 12% de pobres.

Podemos concluir que hubo una ligera reducción de la pobreza para nuestra última muestra, aunque debemos tener en cuenta que no se tuvo en consideración a los individuos que no declararon ingresos, eso explica la notable disminución de registros entre ambos períodos. Tampoco se aplicó ningún ponderador para este cálculo a diferencia de como lo suele hacer INDEC.