



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



## **Maestría en Economía Aplicada**

**Materia:**  
**Taller de Programación**

**Docente:**  
**Noelia Romero**

### **Informe – Trabajo Práctico # 1**

**Primer encuentro con la EPH**

**Grupo # 5:**

**Cammisi, Andrés  
Porco, Matías  
Pineda, David**

**Link GitHub: [https://github.com/DavidPT0902/Grupo\\_5\\_UBA\\_2025.git](https://github.com/DavidPT0902/Grupo_5_UBA_2025.git)**

**Octubre, 2025**

## **PARTE I. FAMILIRIANZANDONOS CON LA BASE EPH Y LIMPIEZA**

La primera parte del trabajo consistió en revisar y preparar las bases de microdatos de la Encuesta Permanente de Hogares (EPH) del Instituto Nacional de Estadística y Censos (INDEC) de Argentina correspondientes al primer trimestre de 2005 y 2025.

Las bases fueron limitadas a la Región Pampeana (43) y contemplan 15 variables de interés comunes en ambos años: 1) codusu, 2) número de hogar (nro\_hogar), 3) sexo (ch04), 4) años cumplidos (ch06), 5) estado civil (ch07), 6) cobertura médica (ch08), 7) nivel educativo (nivel\_ed), 8) condición de actividad (estado), 9) categoría de inactividad (cat\_inac), 10) ingreso per cápita familiar (ipcf), 11) categoría ocupacional (cat\_ocup), 12) cantidad de ocupaciones (pp03d), 13) total de horas que trabajó en la semana en la ocupación principal (pp3e\_tot), 14) total de horas que trabajó en la semana en otras ocupaciones (pp3f\_tot), 15) tipo de empresa o negocio (pp04a), 16) tamaño del establecimiento (pp04c), 17) tiempo buscando trabajo (pp10a), y finalmente, los factores de expansión pondera y pondih.

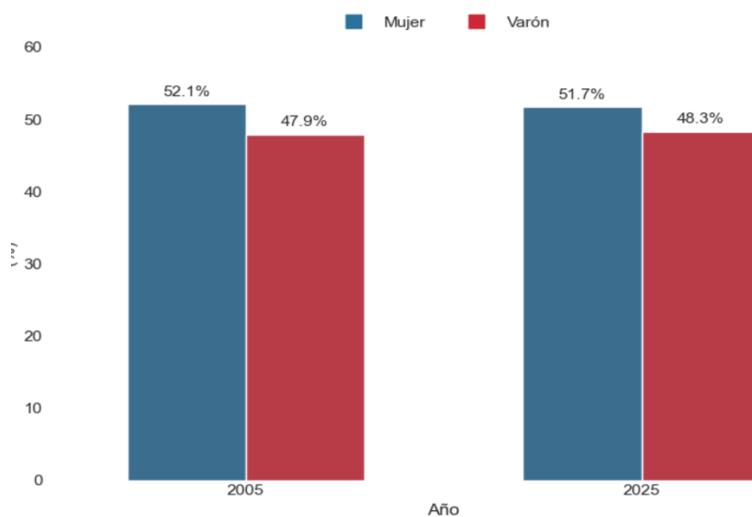
Se procedió a un diagnóstico inicial de valores faltantes por medio de mapas de calor (heatmaps). En 2005, las variables con mayor porcentaje de datos ausentes fueron las horas trabajadas en la ocupación principal y en otras ocupaciones. Por su parte, en el 2025 las variables que presentaron este problema fueron cantidad de ocupaciones, total de horas que trabajó en la semana en la ocupación principal, total de horas que trabajó en la semana en otras ocupaciones, tipo de empresa o negocio, tamaño del establecimiento y tiempo buscando trabajo. Esto podría explicarse porque únicamente contienen información de la población ocupada o asalariada.

La limpieza de las bases de datos se orientó por los criterios establecidos en los diseños de registro y estructura para las bases de microdatos de la EPH. Los valores negativos de la variable edad (ch06) se convirtieron a cero (0). En la variable horas trabajadas a la semana se convirtieron a cero (0) los valores negativos y aquellos superiores a 120 se codificaron como NaN. Para las variables horas trabajadas en la ocupación principal y en otras ocupaciones se recodificaron los valores 99 y 999 como NaN. Los ingresos per cápita familiar negativos se establecieron en NaN. Este procedimiento se realizó en ambas bases. Finalmente, se consolidó una base (eph\_panel\_05\_25.csv), la que contenía 28,454 observaciones: 14,651 en 2005 y 13,803 en 2025.

## PARTE II. PRIMER ANÁLISIS EXPLORATORIO

El análisis exploratorio permitió caracterizar rasgos básicos de la población en la región pampeana utilizando la base consolidada de la EPH para 2005 y 2025. Primero, se comparó la composición por sexo, evidenciando una mayor representación de las mujeres en ambos años. En 2005 comprendían el 52.1% y los varones el 47.9%. Para el 2025 no se observaron cambios relevantes con una distribución de 51.7% de mujeres y 48.3% de varones. Estos datos indican que la región no experimentó cambios demográficos significativos en este período.

Gráfico 1. Composición por sexo — Región Pampeana (2005 y 2025)



En segundo lugar, se estimaron matrices de correlación para conocer las relaciones bivariadas entre variables sociodemográficas y laborales. Entre estas destacan las siguientes: edad, sexo, estado civil, cobertura médica, nivel educativo, condición de actividad, categoría de inactividad e ingreso per cápita familiar. Para realizar este ejercicio se crearon variables *dummies* para representar con mayor consistencia las diferentes categorías en cada año. Las correlaciones se calcularon con el coeficiente de Pearson y se representaron por medio de mapas de calor, en los cuales las tonalidades rojizas indicaban correlaciones positivas y las azuladas correlaciones negativas.

Gráfico 2. Correlación — 2005 (Región Pampeana)

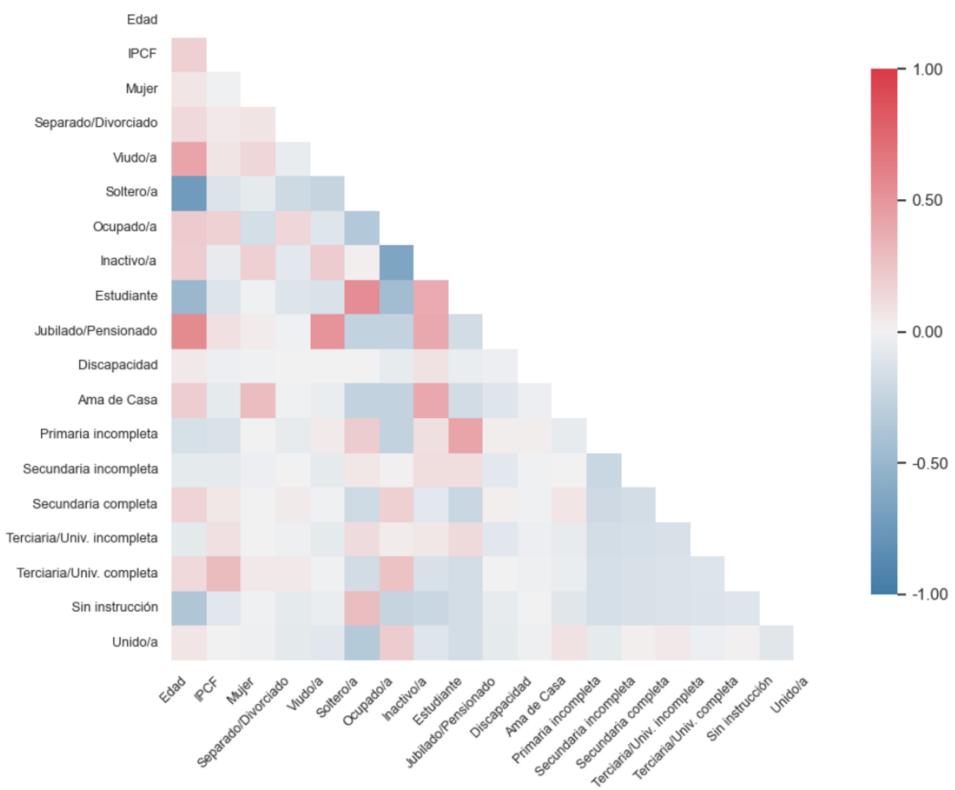
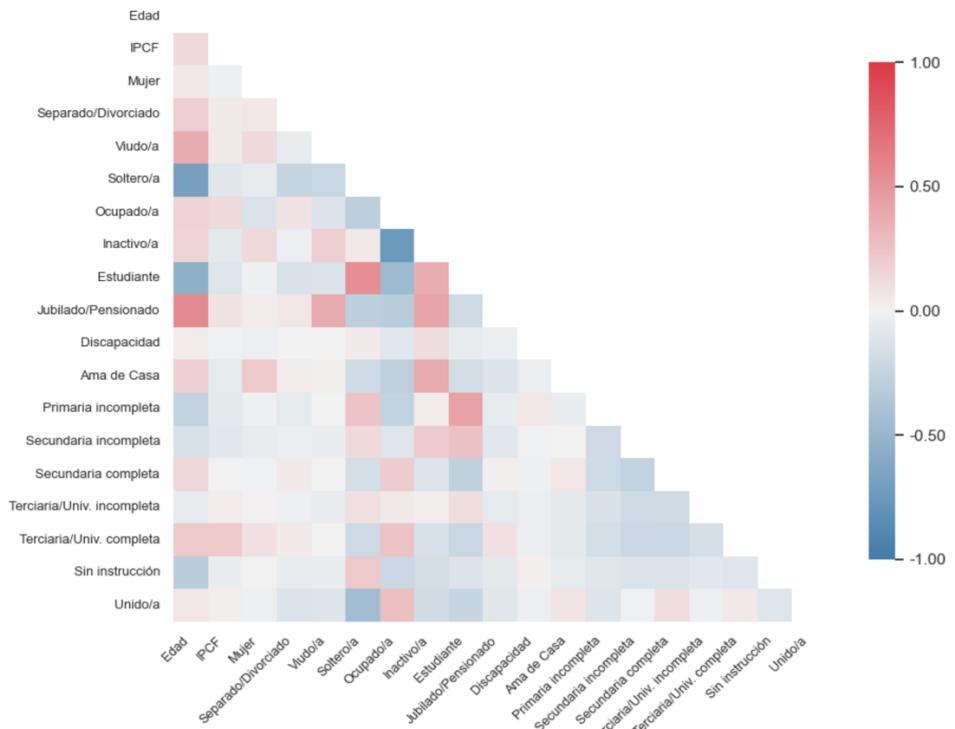


Gráfico 3. Correlación — 2025 (Región Pampeana)



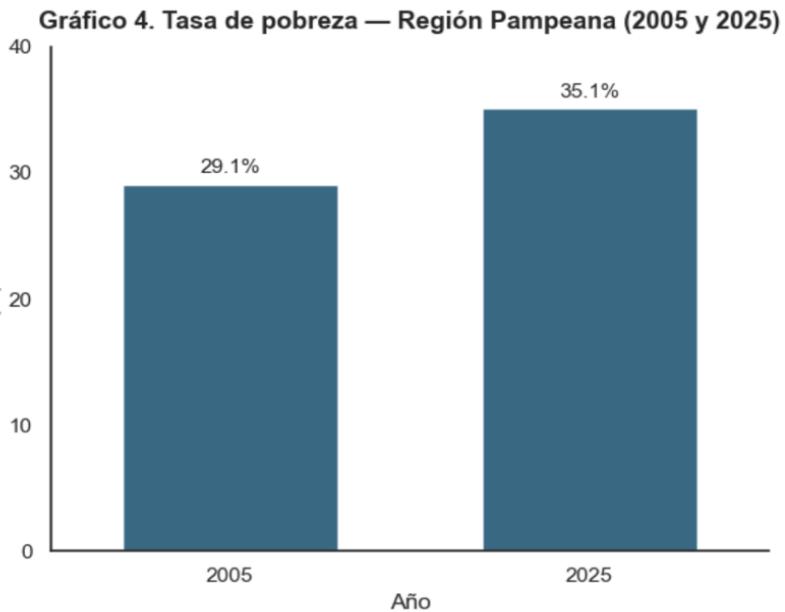
Para el 2005, la edad se correlacionaba positivamente con la probabilidad de estar jubilado o pensionado y negativamente con ser estudiante o inactivo. De igual forma, a mayor nivel educativo se observó una mayor correlación positiva con el ingreso per cápita familiar. Por su parte, en 2025 no se observaron modificaciones evidentes en las correlaciones entre las variables, no obstante, parecen mostrar mayor dispersión en las intensidades, lo que podría indicar algunos cambios en la configuración ocupacional o educativa de la región pampeana.

### **PARTE III. CONOCIENDO A LOS POBRES Y NO POBRES**

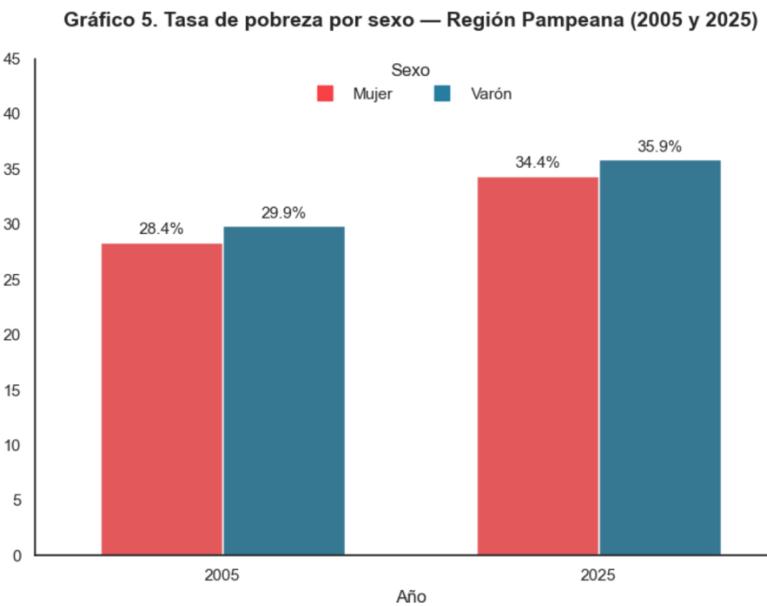
La última sección del trabajo se dedica a explorar la pobreza monetaria de la región Pampeana. En primer lugar, se identificaron los casos con información válida sobre ingresos. Del total de la población, 0.12% no respondieron su condición de actividad y el 16.3% no indicaron su ingreso total familiar; la mayoría de estos casos se contrarion en 2025. Por medio de la tabla\_adult\_equiv.xlsx se agregó a la base de datos consolidada una columna de adultos equivalentes por edad y sexo. Luego se estimó el tamaño equivalente de cada hogar para continuar con el cálculo del ingreso mínimo necesario para que una familia se contabilice como no pobre.

La estimación de la pobreza se realizó comparando el Ingreso Total Familiar de cada hogar con el ingreso mínimo necesario para no ser pobre, calculado como Canasta Básica Total (CBT) por adulto equivalente multiplicada por la cantidad de adultos equivalentes del hogar. Para 2005 el producto de la CBT fue de \$205,07 y para 2025 de \$365.177.

El siguiente paso fue categorizar como pobres las personas de hogares con ITF inferior al ingreso necesario y al resto se les clasificó como no pobres. Como se indica en el Gráfico 4, la proporción de pobres creció en la región Pampeana desde el 29.1% en 2005 al 35.1% en 2025, mostrando un aumento de 6 puntos porcentuales. Además, se constató un incremento en la dispersión de los datos: para el primer año la desviación estándar era de 0.454 y para el segundo de 0.477; lo que sugiere mayor heterogeneidad en la distribución de los ingresos entre los hogares.



Por su parte, en el Gráfico 5 se evidencia que para ambos años la tasa de pobreza fue mayor en varones que en mujeres. En 2005 el 28.4% de las mujeres eran pobres y el 29.9% de los varones; para el 2025 las cifras ascendieron a 34.4% y 35.9%, respectivamente.



Estos datos deben tomarse con cautela, debido a que no contemplan ajustes por inflación. No obstante, el trabajo permite confirmar que la pobreza es un problema persistente en una región importante del país, en la cual casi un tercio de la población no alcanza ingresos mínimos necesarios para cubrir necesidades básicas.