



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



Taller de Programación

Maestría en Economía Aplicada

Tercer trimestre 2025

Trabajo Práctico N° 1

Link GitHub <https://github.com/jorge-ux272/Big-Data-UBA--Grupo-8>

Docente: Noelia Romero

Alumnos:

Jorge Eduardo Bolaños Gamarra

Mario Antonio Valdivia Reyes

Héctor Sebastián San Martín

Introducción

La Encuesta Permanente de Hogares (EPH) es un programa nacional desarrollado por el Instituto Nacional de Estadística y Censos (INDEC), junto con las direcciones provinciales de estadística, para relevar información sobre las características demográficas y socioeconómicas de la población (INDEC, 2025b). Permite recopilar información exhaustiva sobre aspectos como los ingresos, situación laboral, desigualdad, brechas, etc., con lo cual brinda una radiografía sobre el bienestar económico y la distribución de recursos en la población urbana de Argentina.

El presente documento está estructurado en tres partes. En la primera se describe el trabajo de preparación de los datos, selección de variables y limpieza, haciendo también mención a la identificación de las personas pobres bajo la metodología de INDEC. En la segunda parte se hace un análisis exploratorio de los datos, considerando las variables relevantes. En la tercera parte se hace la identificación de los hogares y personas pobres, generando variables y visualizaciones adecuadas.

1. Limpieza base Encuesta Permanente de Hogares (EPH)

El INDEC identifica a las personas pobres en Argentina principalmente a través de la medición monetaria, que compara los ingresos de los hogares con el costo de la Canasta Básica Total (CBT). Se considera que un hogar es pobre si su ingreso total no cubre la CBT, que incluye alimentos, indumentaria, transporte, vivienda, y otros servicios básicos.

En esta investigación, estamos utilizando las bases de datos de la Encuesta Permanente de Hogares (EPH) que lleva a cabo el Instituto Nacional de Estadística y Censos (INDEC) de Argentina (INDEC, 2025a), la cual permite conocer las características sociodemográficas y socioeconómicas de la población. En este caso, utilizamos los periodos correspondientes al primer trimestre de los años 2005 y 2025.

Elegimos para el análisis la región Pampeana, la más poblada del país, en la que se incluyen los centros urbanos con mayor relevancia en términos poblacionales. Las variables seleccionadas para analizar son las que se menciona a continuación:

Tabla 1. Variables seleccionadas para el análisis

Variable	Descripción
CODUSU	Código para distinguir VIVIENDAS
NRO_HOGAR	Código para distinguir HOGARES
COMPONENTE	Nº de orden que se asigna a las personas que conforman cada hogar de la vivienda
CH04	Sexo del miembro del hogar (1 = varón, 2 = mujer)
CH06	Edad
CH07	Estado Civil
CH08	Posee cobertura médica
NIVEL_ED	Nivel educativo más alto alcanzado (sin considerar si está completo o no)
ESTADO	Condición de actividad: ocupado, desocupado, inactivo
CAT_INAC	Categoría dentro de los inactivos (jubilado/a, estudiante, ama de casa, etc.)
ITF	Monto de ingreso total familiar
IPCF	Ingreso per cápita familiar (en pesos corrientes del trimestre)
P21	Monto Ingreso laboral de la ocupación principal
P47T	Monto Ingreso total individual (suma de todos los ingresos personales)
PP3E_TOT	Horas totales trabajadas en la ocupación principal
PP04A	Categoría ocupacional (patrón, cuenta propia, asalariado, etc.)
PP03D	Cantidad de ocupaciones
DECIFR	Nº de decil de ingreso total familiar del total EPH
CH09	Sabe leer o escribir
REGIÓN	Región geográfica (PAMPEANA)
ANIO	Año del relevamiento

Nota. Elaboración propia, 2025.

Es importante comentar la cantidad de observaciones que tenemos para cada año: de las 47.030 observaciones de 2005, 14.651 de ellas corresponden a Pampeana, mientras que de las 45.425 observaciones de 2025, solo 4.746 se refieren a dicha región. Eso quiere decir que en este análisis estamos contando con 19.397 observaciones entre los trimestres de ambos años.

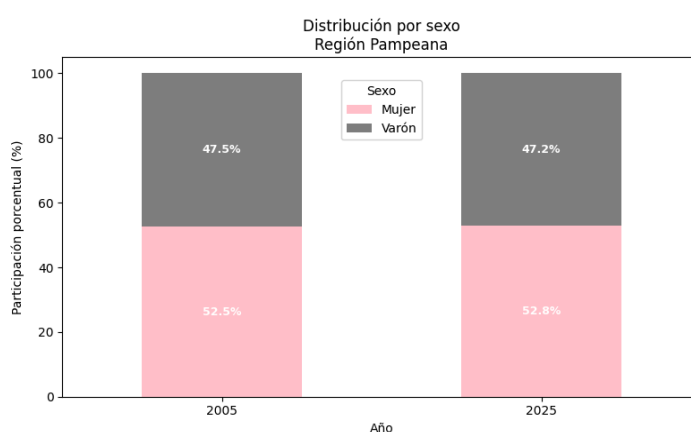
Vale también destacar que en la base de 2005 no se encontraron valores *missing*, pero sí en la de 2025 (8.457 valores *missing*).

Las siguientes operaciones se realizaron para homogeneizar el formato de los datos:

1. Se ajustaron los valores y los formatos de la base de 2005 para homologarlos a la de 2025
2. Para el caso de la edad (CH06) en la base de 2025, se agruparon a las personas de 98 año o más en un mismo grupo para homologar con la base de 2005
3. Los valores -1, -2, -3 y -9 se registraron como NA
4. En las columnas de CAT_INAC y PP04A se ajustan los valores 0 a NA, ya que en el manual del EPH no se indica que éstas puedan adoptar valor 0.

2. Primer Análisis Exploratorio

Figura 1. Participación por sexo en la Región Pampeana (primer trimestre).



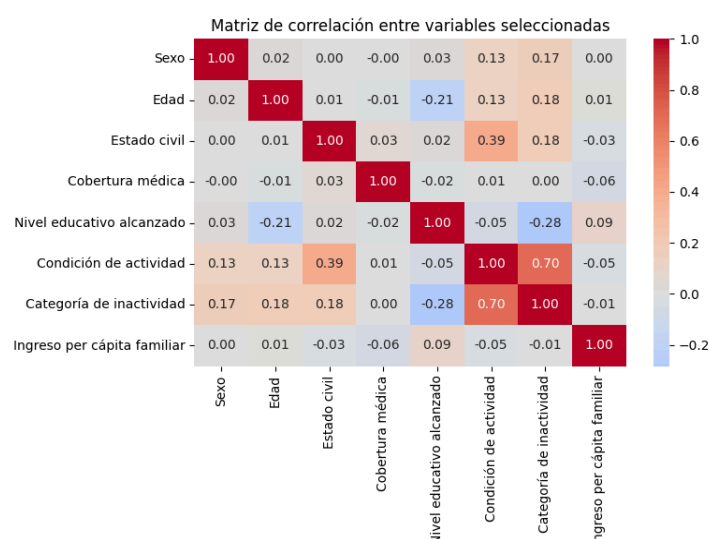
Nota. Elaboración propia con datos de la EPH 2025.

El gráfico muestra la distribución porcentual de la población por sexo para los años 2005 y 2025 en la Región Pampeana, según los datos de la Encuesta Permanente de Hogares (EPH) correspondientes al primer trimestre de cada año. En ambos períodos se observa una estructura demográfica estable en términos de composición por sexo. Las mujeres representan aproximadamente el 52 % de la población, mientras que los varones conforman el 48 % restante, manteniéndose prácticamente sin cambios a lo largo del tiempo.

Esta estabilidad refleja que, a pesar de las variaciones en la cantidad total de personas registradas entre ambos años, la proporción relativa entre mujeres y varones se mantiene constante. En línea con las tendencias demográficas nacionales, puede advertirse una ligera mayor participación femenina, coherente con la mayor esperanza de vida de las mujeres y la evolución poblacional observada en las proyecciones del país. En síntesis, el

gráfico evidencia que no se produjeron cambios significativos en la composición por sexo dentro de la Región Pampeana durante el período analizado.

Figura 2. Matriz de correlación de variables de interés



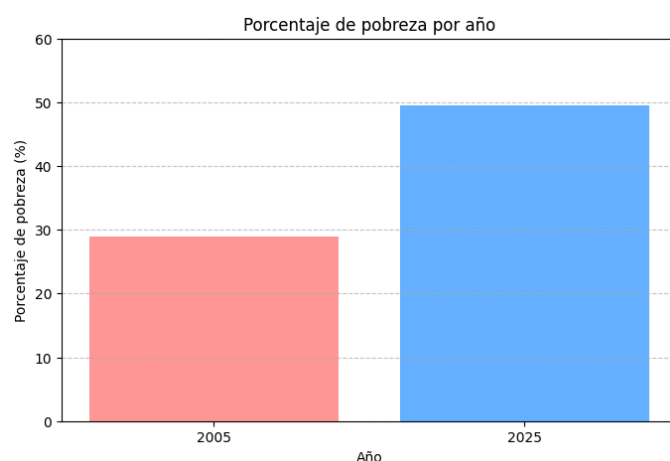
Se observa una correlación positiva moderada entre el nivel educativo alcanzado y el ingreso per cápita familiar, dado que la mayor parte de los valores están por debajo de 0,4. Esto sugiere que mayores niveles de instrucción están asociados con mejores condiciones económicas.

Asimismo, la edad muestra una correlación positiva con la categoría de inactividad (0,18), reflejando que los grupos etarios mayores tienden a estar fuera del

mercado laboral, posiblemente por jubilación. Por otro lado, la condición de actividad presenta una correlación negativa con la categoría de inactividad, como era de suponerse, dado que ambas variables capturan estados mutuamente excluyentes. Las correlaciones entre sexo y otras variables son débiles, lo que indica que, en este conjunto de datos, el género no está fuertemente asociado con diferencias en educación, ingreso o actividad.

3. Conociendo a los pobres y no pobres

Figura 3. Porcentaje de pobreza en la Región Pampeana (primer trimestre)



De los encuestados en la región Pampeana, 24 personas no respondieron su condición de actividad tanto en 2005 como en 2025, pero en el primer año representaba 0.16% del total, y en 2025 eran el 0.51%. Por lo tanto, en veinte años se registró un aumento en términos relativos de la cantidad de personas que no responden dicha cuestión en 0.35 puntos porcentuales, casi el doble del punto inicial.

Dado que, para el primer trimestre del año 2005, la Canasta Básica Total para un adulto equivalía aproximadamente a ARS 205,07, se identificaron 4.195 personas que no cubrían dicho valor, lo que representa el 28,7% de la muestra. En cambio, para el respectivo trimestre del año 2025, se identificaron 1.652 personas que no cubrieron los ARS 365.177 de la Canasta Básica Total, que representan el 49.46% de la muestra.

Lo anterior refleja un desmejoramiento generalizado de los niveles de pobreza en la región Pampeana: hace veinte años, casi tres personas de cada diez cuyos hogares reportaron su ingreso total familiar en la EPH eran consideradas pobres, mientras que ahora, en el primer trimestre de 2025, fueron casi cinco.

Figura 4. Nivel educativo de personas pobres y no pobres (primer trimestre)

Año	Nivel educativo	No pobre	Pobre
2005	Primario completo	66.95%	33.05%
	Primario incompleto (incluye educación especial)	56.14%	43.86%
	Secundario completo	82.66%	17.34%
	Secundario incompleto	65.43%	34.57%
	Sin instrucción	60.00%	40.00%
	Superior y universitario completo	95.49%	4.51%
	Superior y universitario incompleto	86.88%	13.12%
2025	Primario completo	53.52%	46.48%
	Primario incompleto (incluye educación especial)	38.37%	61.63%
	Secundario completo	52.64%	47.36%
	Secundario incompleto	34.24%	65.76%
	Sin instrucción	37.39%	62.61%
	Superior y universitario completo	82.25%	17.75%
	Superior y universitario incompleto	58.36%	41.64%

Se puede concluir que, tanto en 2005 como en 2025, a medida que aumenta el nivel educativo la proporción de personas pobres disminuye significativamente, confirmando la relación inversa entre educación y pobreza. La educación superior continúa actuando como un factor protector, ya que quienes poseen estudios terciarios o universitarios siguen siendo el grupo con menor incidencia de pobreza. No obstante, su capacidad de protección se redujo: mientras en 2005 apenas un 4,5% de las personas con educación universitaria completa eran pobres, en 2025 esa proporción asciende a casi un 18%.

Referencias bibliográficas

Instituto Nacional de Estadística y Censos (INDEC). (2025a, agosto). *Encuesta Permanente de Hogares: Diseño de registros y estructura para las bases preliminares Hogar y Personas*. INDEC. <https://www.indec.gob.ar>

Instituto Nacional de Estadística y Censos (INDEC). (2025b). *Encuesta Permanente de Hogares (EPH) total urbano: Evolución de la distribución del ingreso, tercer trimestre de 2024*. Informes técnicos, 9(39). <https://www.indec.gob.ar>