



UBA FCE
Universidad de Buenos Aires
Facultad de Ciencias Económicas

Taller de Programación – Trabajo Práctico N.º 2 (2025)

Tema: Métodos No Supervisados y Visualización

Grupo 4

Ángel Enrique Zapata Barros

David Andrés Robalino Chica

Federico Walter Kiswa

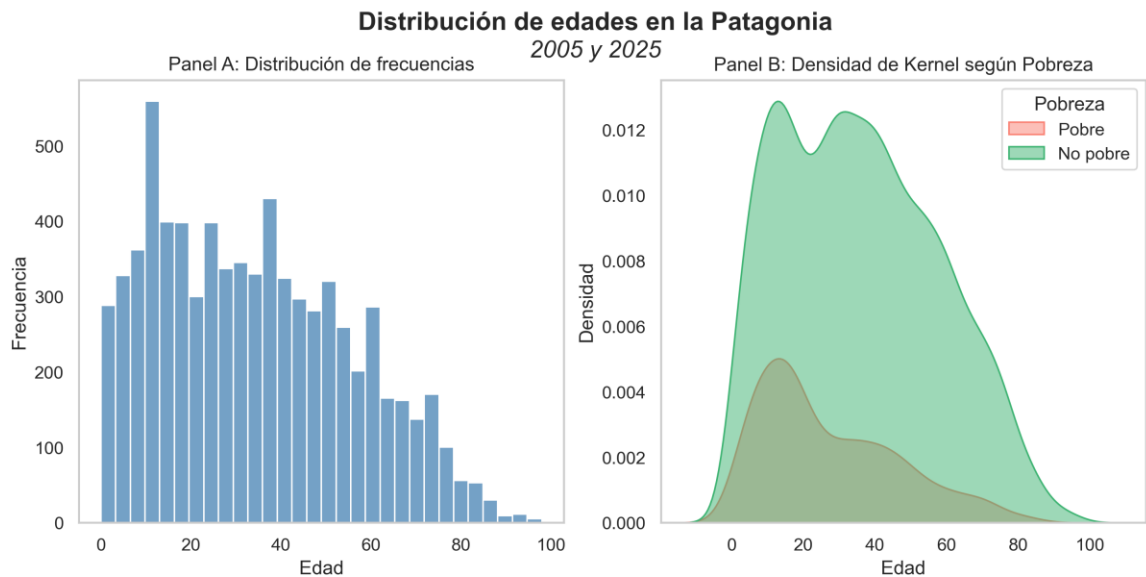
Docente: María Noelia Romero

Repositorio GitHub del Grupo:

https://github.com/Analisis1983/Grupo4_UBA_2025

Parte I: Creación de variables, histogramas, kernels y resumen de la base de datos final

1) Distribución de edades en la Patagonia



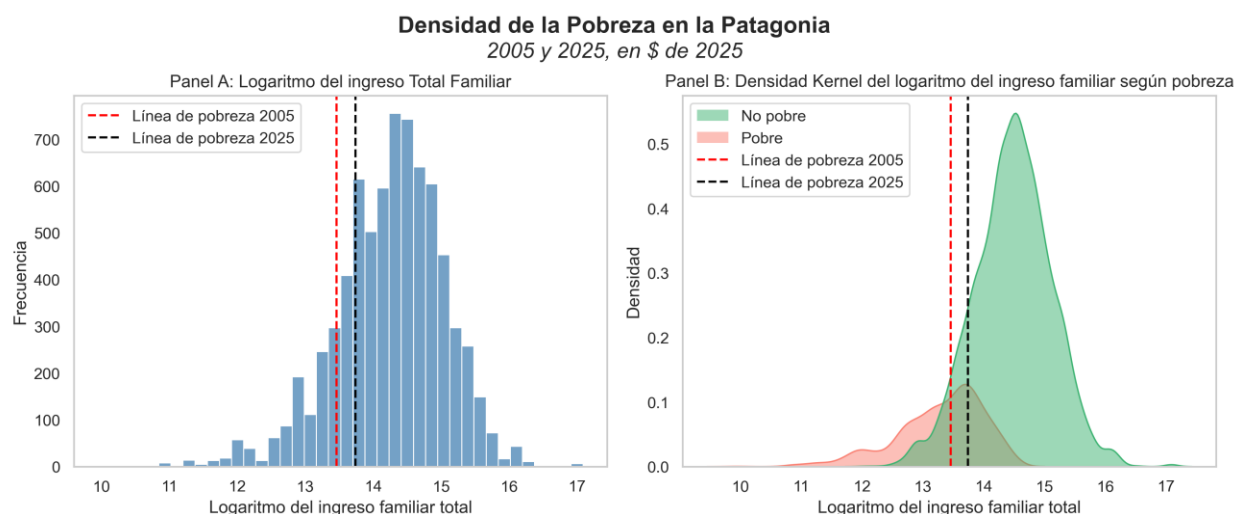
Fuente: Elaboración propia en base a EPH-INDEC

La distribución de la edad presenta una concentración mayor en los grupos jóvenes y adultos hasta los 40 años, con una disminución progresiva en edades avanzadas. La densidad por condición de pobreza muestra que las personas en situación de pobreza tienden a centrarse en edades menores, mientras que los no pobres presentan una distribución más amplia. Esta diferencia sugiere contrastes demográficos vinculados a la inserción laboral y a la estructura etaria de los hogares.

2) Análisis de los años de educación en la Patagonia

El promedio de observaciones cursó 9 años y medio de educación. El máximo cursó 23 años que es quien cursó posgrado y lo finalizó. En promedio, el número de años de educación de las personas varía [5,65] años respecto al promedio general (9,5 años). La mediana es 11, es decir, la distribución está sesgada hacia la izquierda por haber un número importante de personas con baja cantidad de años que tira el promedio para abajo. Aumentó la media de educación en 3 años adicionales. El promedio era del 7,76 y pasó a 10,83

3) Análisis del ingreso familiar



Fuente: Elaboración propia en base a EPH-INDEC

La comparación muestra un corrimiento hacia la derecha de la línea de pobreza en 2025, reflejando el incremento en el umbral monetario necesario para superar la condición de pobreza.. Los hogares pobres (área roja) se concentran claramente por debajo del umbral, mientras que los no pobres (área verde) presentan una distribución desplazada hacia niveles de ingreso más altos. La superposición parcial entre ambas curvas sugiere la existencia de un grupo significativo de hogares cercanos a la línea de pobreza, cuya situación podría variar ante cambios marginales en los ingresos o en los precios de la canasta básica.

4) Análisis de las horas trabajadas por los jefes de hogar.

El promedio de las horas trabajadas en una semana fue de 15,61, con un máximo de 84 horas (12 horas por día). Hubo un aumento del promedio de horas trabajadas en 2025 respecto a 2005, al pasar de 15,22 a 15,90.

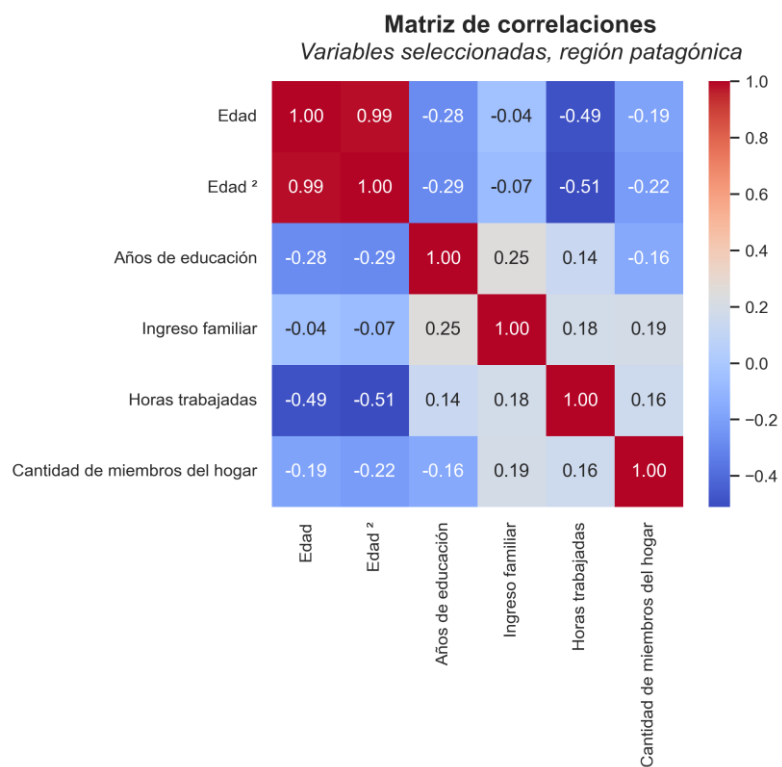
5) Tamaño de la base de datos

	2005	2025	Total
Cantidad observaciones	3206	4134	7340
Cantidad de observaciones con NAs en la variable “Pobre”	0	0	0
Cantidad de Pobres	575	850	1425
Cantidad de No Pobres	2631	3284	5915
Cantidad de variables limpias y homogeneizadas	35	35	35

Parte II: Métodos no supervisados

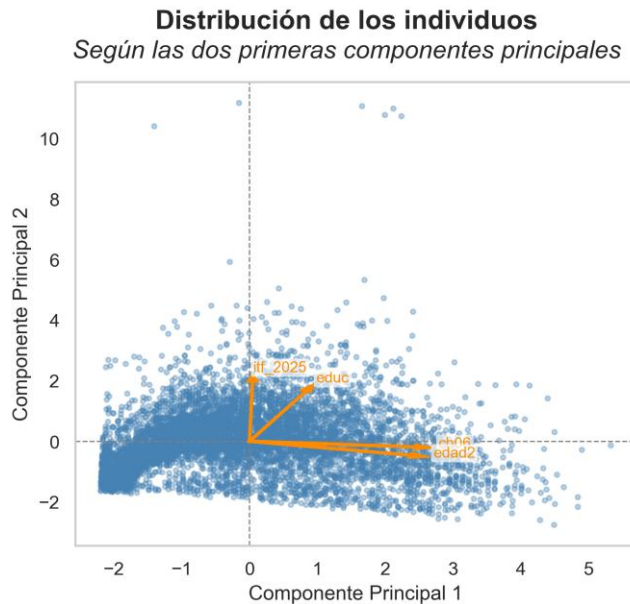
1) Análisis de correlaciones entre variables.

La matriz de correlaciones arroja una de las correlaciones negativas más fuertes entre la cantidad de horas trabajadas y la edad, lo que resulta razonable dada la presencia de personas jubiladas en los grupos etarios más elevados. También, se observa una correlación positiva relevante entre ingreso familiar y años de educación. También se observa, llamativamente, que hay correlación negativa entre años de educación y edad, lo que podría reflejar que los jefes de hogar más grandes provienen de una generación en la los niveles de educación en la sociedad eran menores.

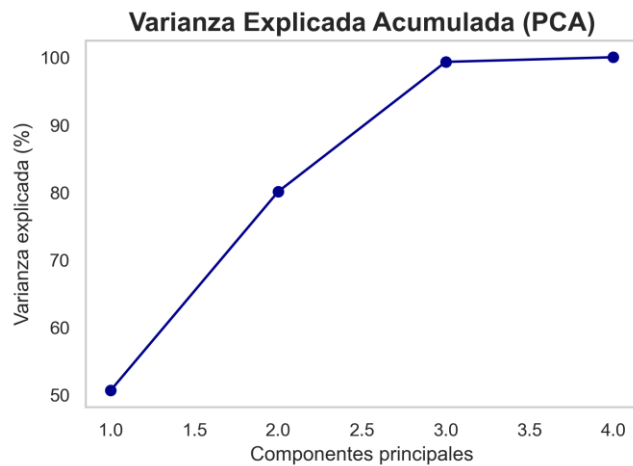


2) Gráfico PCA con ingresos

La mayor dispersión sobre el primer componente sugiere que las diferencias socioeconómicas entre los individuos de la región patagónica se explican principalmente por una dimensión asociada a variables como el ingreso y la educación. La concentración central indica similitud en gran parte de la población, posiblemente vinculada a estructuras laborales y niveles de ingreso intermedios. La presencia de puntos más alejados refleja heterogeneidad interna, típica de economías regionales con contrastes entre áreas urbanas y rurales o entre sectores productivos de distinta intensidad económica.



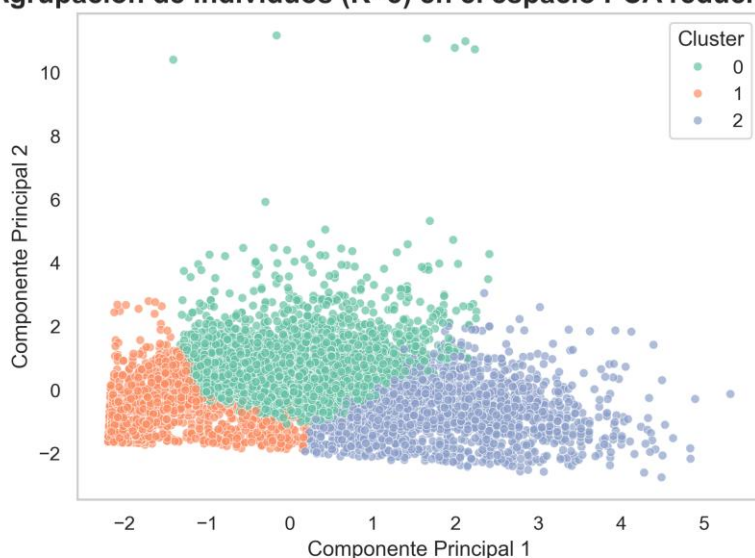
3) Gráfico de varianza explicada.



La varianza explicada acumulada indica que los tres primeros componentes concentran casi la totalidad de la variabilidad del conjunto de datos, lo que sugiere una estructura bien representada con pocas dimensiones. En particular, los dos primeros componentes ya capturan una proporción sustancial de la información, por lo que el modelo puede simplificarse sin pérdida significativa de contenido analítico.

4) Gráfico de clusster de k-medias

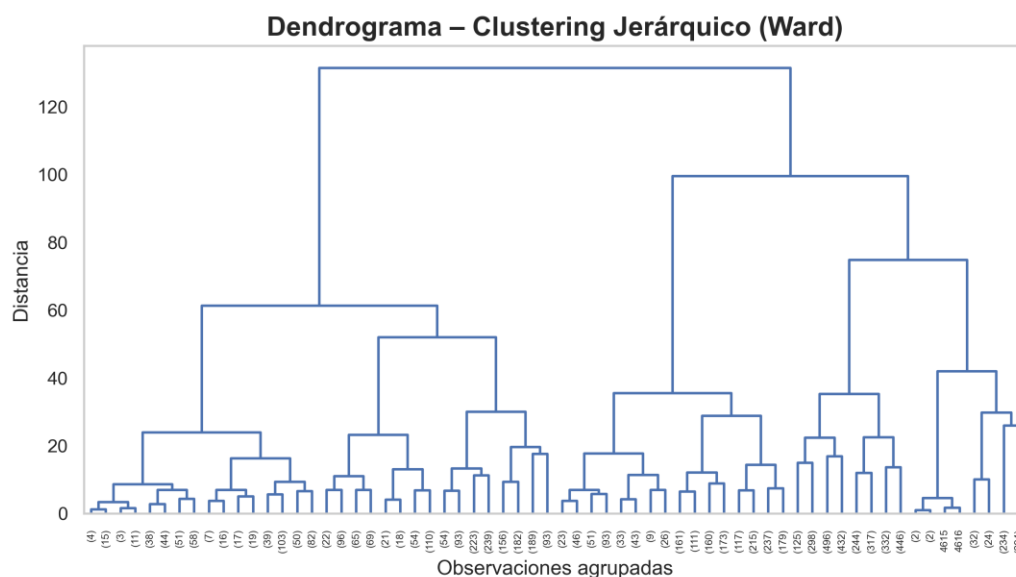
Agrupación de individuos (K=3) en el espacio PCA reducido (2D)



Fuente: Elaboración propia en base a EPH-INDEC

El agrupamiento distingue tres perfiles socioeconómicos: uno de bajos ingresos y educación, otro intermedio y un tercero con mejores condiciones. Esta segmentación resume la desigualdad estructural de la región y la coexistencia de grupos con niveles claramente diferenciados de bienestar.

6) Clúster jerárquico



Fuente: Elaboración propia en base a EPH-INDEC

El dendrograma sugiere una estructura regional heterogénea, con al menos tres grupos socioeconómicos diferenciados por nivel educativo, edad e ingreso. Las distancias de fusión más amplias reflejan brechas

marcadas entre segmentos de la población, posiblemente asociadas a desigualdades en el acceso a educación y oportunidades laborales. Esta configuración evidencia que, dentro del espacio patagónico analizado, coexisten perfiles con características socioeconómicas contrastantes que explican parte de la dispersión observada en los indicadores de bienestar.