Imagen que contiene Texto

Descripción generada automáticamente

Universidad de Buenos Aires.

Facultad de Ciencias Económicas

Escuela de Estudios de Posgrado.

**MAESTRÍA EN ECONOMÍA APLICADA**

Taller de Programación

Segundo Trabajo Práctico:

**Histogramas, Kernels & Métodos No Supervisados usando la EPH**

**Autores:**

Kostzer Federico

Meza Palma Manuel H.

Pardini Maximiliano G.

**docente:** Romero Maria Noelia

Octubre 2025

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

La edad de los individuos en la muestra válida (≥1 año) presenta una media de 35,2 años y una desviación estándar de 22,3 años, con un rango que va de 1 a 97 años. Al desagregar por condición de pobreza, se observa que los individuos no pobres tienen una media de edad de 36,7 años (mínimo 1, máximo 97), mientras que los pobres presentan una media considerablemente menor de 24,3 años (mínimo 1, máximo 90). La mediana de edad es 34 años para los no pobres y 18 años para los pobres. La muestra está compuesta por 20.714 individuos no pobres y 2.838 pobres, representando estos últimos aproximadamente el 12% de la muestra.

Gráfico, Gráfico de líneas, Histograma

El contenido generado por IA puede ser incorrecto.

Los gráficos exploratorios muestran en el Panel A la distribución de edad de toda la población mediante un histograma, evidenciando una mayor concentración de individuos jóvenes y adultos jóvenes. En el Panel B, la distribución kernel por condición de pobreza revela que los pobres se concentran en edades más tempranas, mientras que los no pobres se distribuyen en un rango más amplio de edades, confirmando la diferencia observada en las estadísticas descriptivas.

Se construyó la variable **“educ”**, que representa la cantidad total de años de educación formal alcanzados por cada individuo. Para su elaboración se utilizaron las variables **CH12 (nivel educativo más alto cursado o cursando)**, **CH13 (finalización del nivel)** y **CH14 (último año aprobado)**.  
A partir de estas, se asignaron valores acumulativos de años de escolaridad según el nivel alcanzado y, en los casos en que el nivel no fue completado, se consideraron los años efectivamente cursados.

El promedio de educación formal en la muestra fue de **6,56 años**, con una desviación estándar de **5,07 años**, un valor mínimo de **0 años** y un máximo de **18 años**. El **percentil 50 (mediana)** se ubicó en **6 años**, lo que indica que la mitad de los individuos posee educación primaria completa o inferior.  
La distribución muestra una alta concentración en niveles bajos de educación, con una cola derecha más extensa, lo que refleja la menor proporción de personas con educación Gráfico, Histograma

El contenido generado por IA puede ser incorrecto.terciaria o universitaria.

3.

Gráfico

El contenido generado por IA puede ser incorrecto.

La distribución del ingreso total familiar ajustado a pesos de 2025 es **sesgada a la derecha**, con muchos hogares concentrados en niveles bajos y unos pocos con ingresos muy altos que elevan la media. La mediana es $637.200, mientras que el promedio es $926.771, mostrando que la presencia de ingresos extremos eleva la media respecto a la mediana. El rango intercuartílico va de $322.848 a $1.180.944, reflejando que el 50% central de los hogares se concentra en este intervalo, y el mínimo y máximo son $8.496 y $26.000.000, respectivamente. En los gráficos, los hogares pobres se concentran por debajo de la línea de pobreza calculada en el TP1, mientras que los no pobres se distribuyen en rangos más altos con mayor dispersión.

Coeficiente de actualización 2005–2025 obtenido multiplicando los factores de inflación oficial anual y mensual reportados por el INDEC y complementados con fuentes confiables (lanacion.com.ar, infobae.com, diariodemocracia.com, diarioc.com.ar, mdzol.com, pagina12.com, ámbito.com, chequeado.com).

4.

La variable horastrab, definida como la suma de horas trabajadas en la ocupación principal y otras ocupaciones, muestra un promedio de 15,5 horas semanales, mientras que la mediana es 0, lo que indica que más de la mitad de los jefes de hogar no reportó horas trabajadas en la semana de referencia. El mínimo es 0 y el máximo 1.998, reflejando algunos valores atípicos o posibles errores de registro. El rango intercuartílico va de 0 a 30 horas, mostrando que el 50 % central de los jefes que sí trabaja realiza jornadas relativamente cortas. En conjunto, la distribución evidencia una alta proporción de jefes inactivos o desempleados junto con casos de actividad laboral.

5.

La base final para la región 43 contiene 23.821 observaciones en total, con 14.481 del año 2005 y 9.340 del año 2025. No hay valores faltantes en la variable “pobre”. Se registran 2.900 hogares pobres (1.780 en 2005 y 1.120 en 2025) y 20.921 hogares no pobres (12.701 en 2005 y 8.220 en 2025). La cantidad de variables limpias y homogeneizadas es de 261, manteniendo consistencia en ambos períodos. Esto permite realizar análisis comparativos confiables entre los dos trimestres.

|  |  |  |  |
| --- | --- | --- | --- |
| **Concepto** | **2005** | **2025** | **Total** |
| Cantidad observaciones | 14.481 | 9.340 | 23.821 |
| Cantidad de observaciones con NAs en la variable "pobre" | 0 | 0 | 0 |
| Cantidad de Pobres | 1.780 | 1.120 | 2.900 |
| Cantidad de No Pobres | 12.701 | 8.220 | 20.921 |
| Cantidad de variables limpias y homogeneizadas | 261 | 261 | 261 |

Parte II: Métodos No Supervisados

1.

2.

3.

4.

5.

6.

7.