

TRABAJO PRÁCTICO

PROBABILIDAD Y ESTADÍSTICA

Accurso Agustín, Barbarroja Federico, Stefano Candela

Tecnicatura universitaria en Inteligencia Artificial
FCEIA - UNR

Docentes: Avila Aylén, Gottig Nicolás, L'heureux Julián



INTRODUCCIÓN Y METODOLOGÍA

En el presente informe se analizarán los datos obtenidos en la Encuesta Permanente de Hogares (EPH) del primer trimestre de 2024. La EPH es un programa nacional de producción permanente de indicadores sociales cuyo objetivo es conocer las características sociodemográficas y socioeconómicas de la población. Se realiza en forma conjunta por el INDEC y las Direcciones Provinciales de Estadística.

La EPH es una encuesta que se basa en una muestra probabilística, estratificada, en dos etapas de selección. En la primera, dentro de cada aglomerado, se selecciona una cantidad de áreas. El método por el que se seleccionan es el del muestreo estratificado, con probabilidad de selección proporcional a la cantidad de viviendas particulares del área. En la segunda etapa, se listan todas las viviendas particulares de las áreas seleccionadas, y se seleccionan las viviendas a encuestar mediante el muestreo sistemático de segmentos de viviendas contiguas. Los listados de viviendas se ordenan geográficamente previamente a la selección, lo que constituye una estratificación geográfica implícita.

Si bien esta encuesta se trata de un estudio observacional, donde la población son todas las personas habitantes de Argentina, en el presente análisis se tomará en cuenta a las personas que contestaron en los campos de interés, y tienen entre 18 y 65 años, ya que se busca hacer análisis relacionados a los ingresos totales, y los individuos fuera de ese rango etario, podrían generar ruido al estudio de los parámetros de interés. Entonces, la **unidad observacional** será cada persona encuestada que cumpla con los criterios de inclusión ya mencionados, y la **población estadística**, todos los habitantes de Argentina que cumplan dichos criterios.

Se determinan entonces cuatro variables relevantes con sus parámetros de interés asociados, a partir de la encuesta de la EPH ya mencionada y descrita.

Las **variables** de interés seleccionadas, y su clasificación, son las siguientes:

- **¿Cuál es el nivel más alto que cursa o cursó? (CH12):** Cualitativa Nominal.
- **¿Finalizó ese nivel? (CH13):** Cualitativa Nominal.
- **Monto de ingreso total individual (P47T):** Cuantitativa De Razón.
- **Monto del ingreso por subsidio o ayuda social (V5_M):** Cuantitativa De Razón.

Por otro lado, los **parámetros** asociados, entre otros, serán los siguientes:

- Distribución y proporciones de niveles educativos alcanzados.
- Media y mediana de ingresos según nivel educativo.
- Porcentaje de personas que reciben subsidios, clasificado por nivel alcanzado.
- Promedio de ingresos por subsidios.

En otras palabras, se analizará la relación entre el nivel educativo alcanzado y los ingresos personales, también indagando hasta qué nivel educativo llegan las personas que reciben subsidios o ayudas sociales. Además de estadísticas de centralidad y dispersión para las variables seleccionadas, el análisis, potencialmente, podría mostrar las desigualdades sociales y económicas asociadas a la formación educativa de las personas. Se eligió la variable de ingresos totales y no la de salario, porque esto podría ser de interés para analizarla no sólo como consecuencia del nivel educativo alcanzado, sino también como posible causa.

ANÁLISIS DESCRIPTIVO

Como primera medida, al depurar el DF se visualizó que una gran cantidad de datos no resultaban útiles para los parámetros establecidos, por lo que se filtraron muchas de las respuestas que dió la gente (como fue explicado en la metodología). Se optó entonces, por analizar primero qué porcentaje del total de la encuesta (muestra original) fue tomado ahora como muestra, representado en un gráfico circular.

Distribución: Filtrados vs No Filtrados

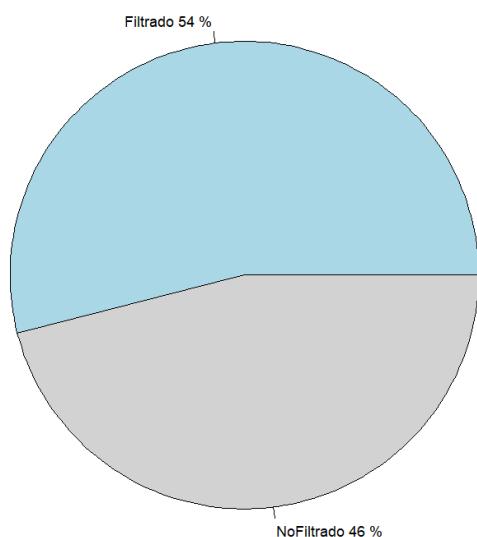


Gráfico 1

Para visualizar la cantidad de personas que alcanzaron cada nivel educativo, se utilizó un gráfico de barras. Los números del 1 al 9 simbolizan los siguientes niveles:

- | | |
|-----------------------|--|
| 1 = Jardín/preescolar | 6 = Terciario |
| 2 = Primario | 7 = Universitario |
| 3 = EGB | 8 = Posgrado universitario |
| 4 = Secundario | 9 = Educación especial (discapacitado) |
| 5 = Polimodal | |

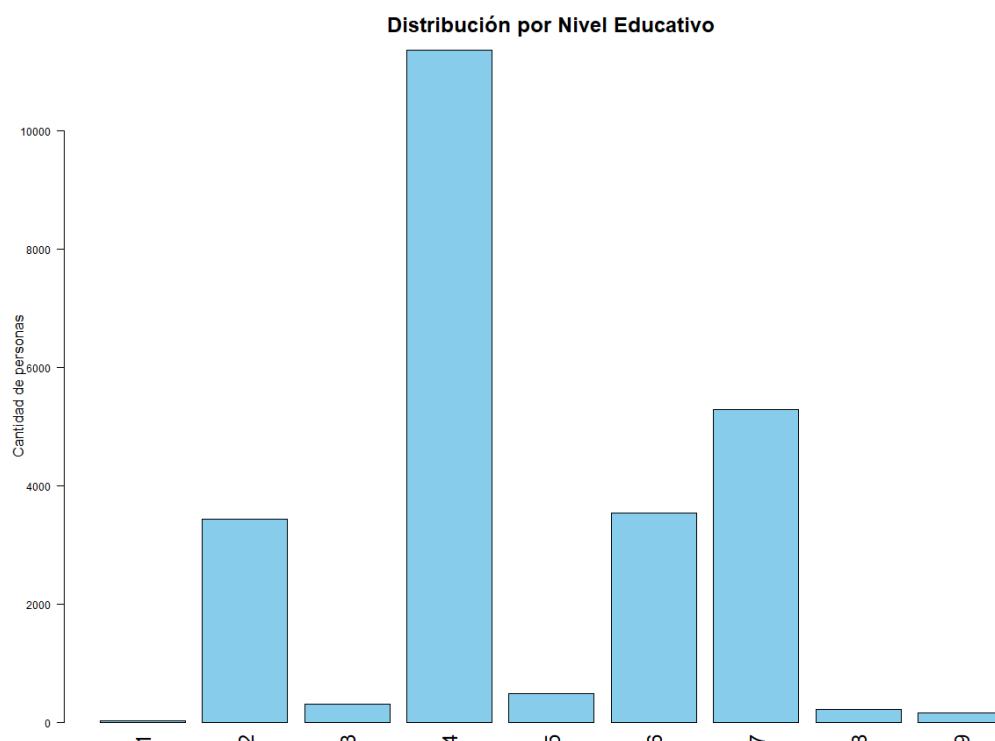


Gráfico 2.

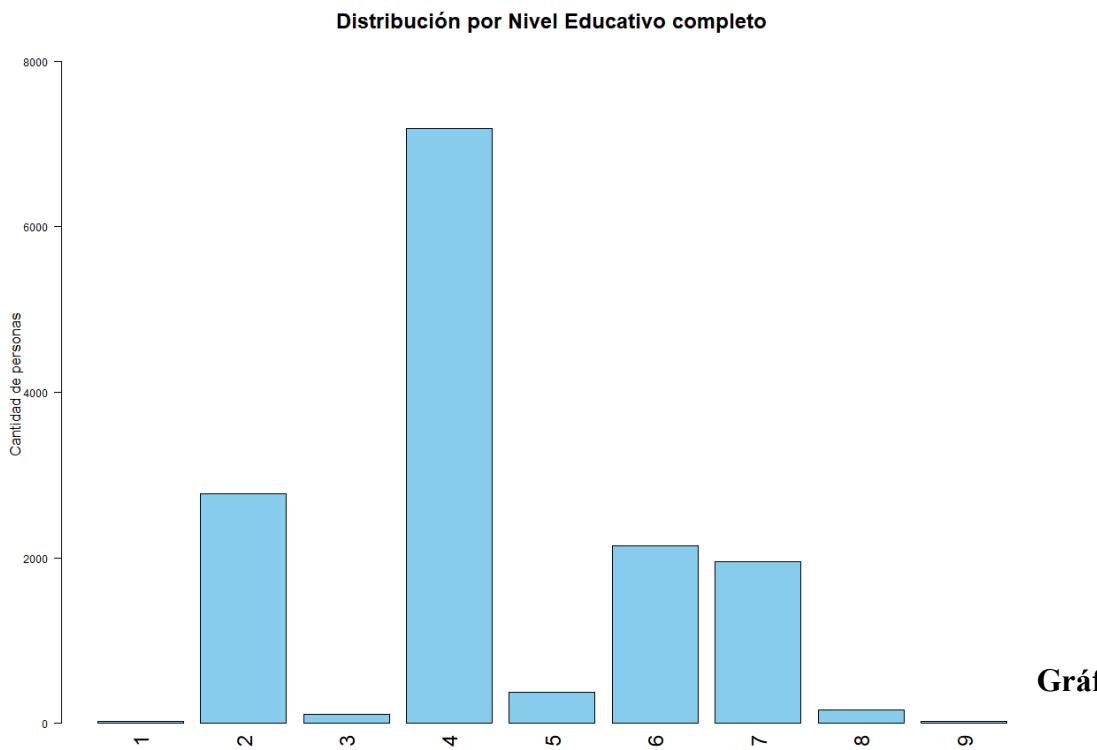


Gráfico 3.

En el gráfico 2, se puede ver que muy pocas personas solo alcanzaron el nivel 1, es decir, jardín/preescolar. El número de registros para dicho valor de esa variable es de 27, un 0.11% del total de registros. En cambio, el nivel educativo alcanzado más registrado es el secundario, con 11366 registros (45.72% del total). Por otro lado, lo que permite visualizar el gráfico 3 es cuántos finalizaron cada nivel, pero principalmente si la distribución se mantiene respecto de quienes lo alcanzaron: podemos notarla similar en todos los niveles, sin embargo, el nivel universitario (7) es el de mayor diferencia.

Se sigue adelante con la variable de ingreso total individual, que no será analizada por sí sola sino en relación con el nivel educativo alcanzado.

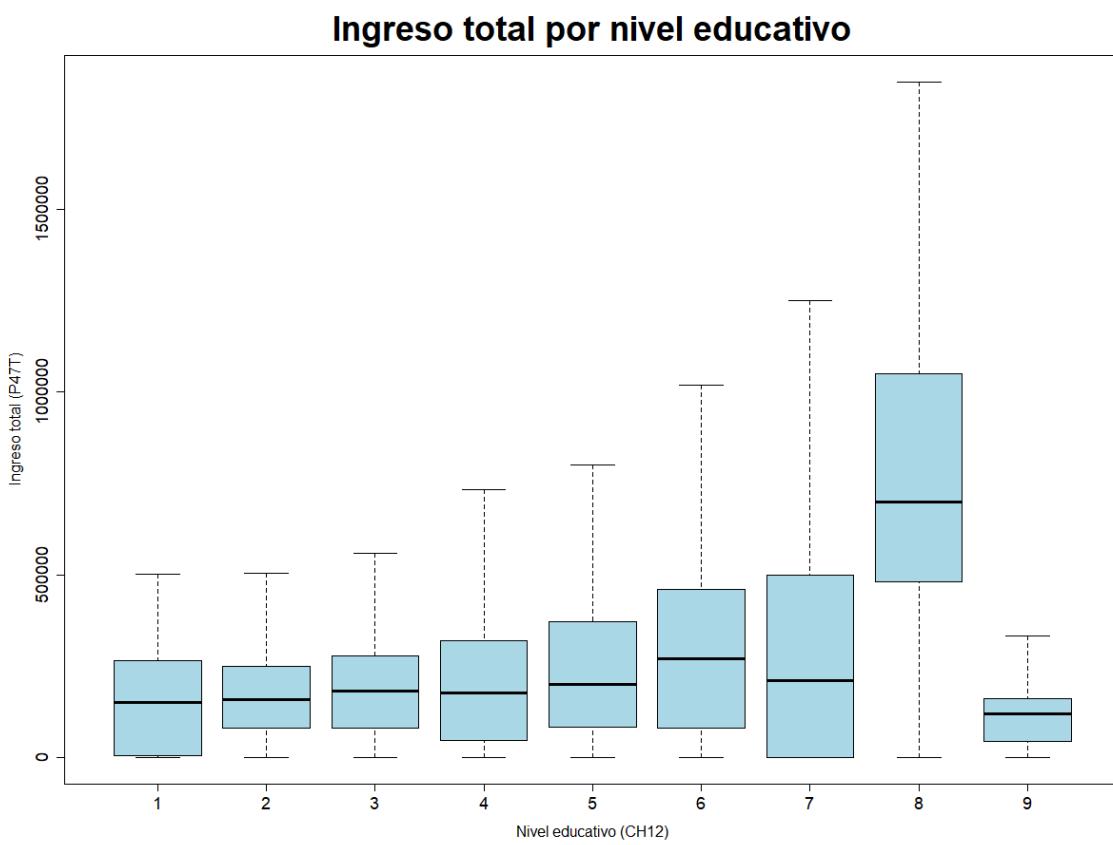
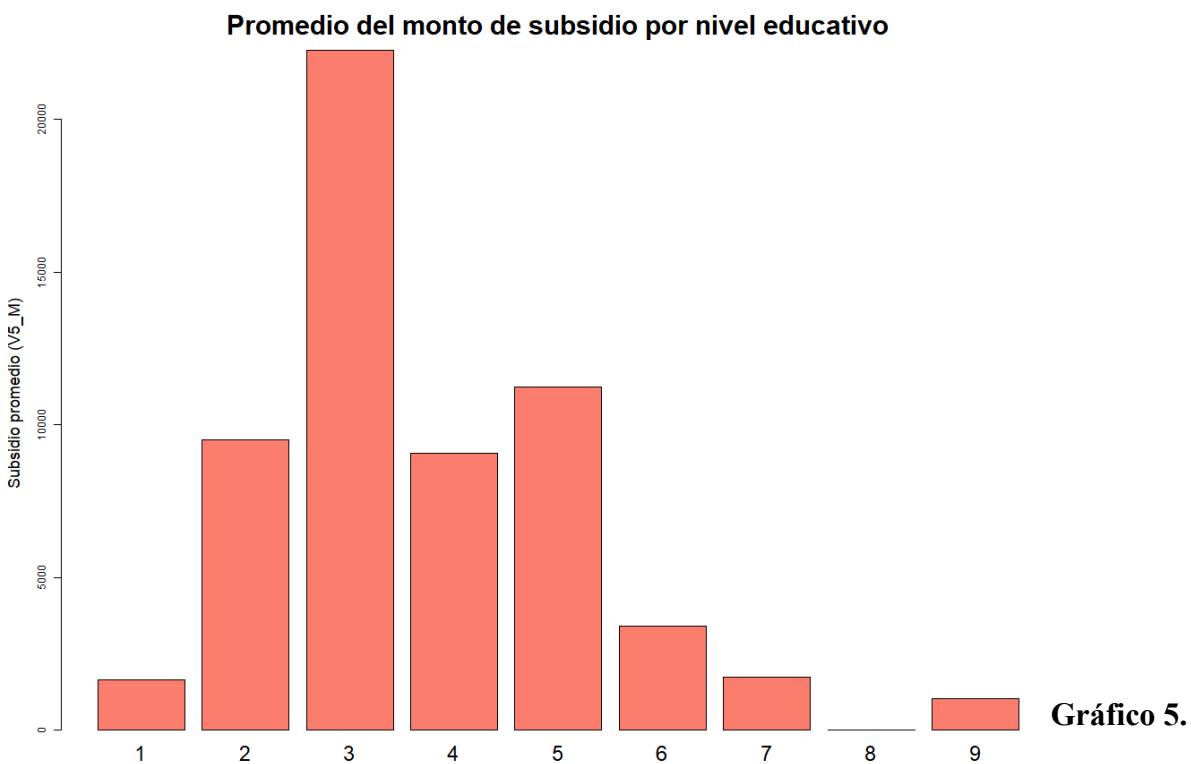


Gráfico 4.

Se observa en el gráfico 4 que las personas con posgrado universitario sobresalen del resto en cuánto a sus ingresos totales. Para el presente diagrama ‘boxplot’ no se tuvieron en cuenta los valores atípicos porque impedían la visualización clara de los datos.

Ahora, en el gráfico 5, se analizará el monto promedio de ingresos por subsidios, también según el nivel educativo alcanzado. Vemos que la mayoría de los subsidiados están en los ciclos primarios y secundarios (incluyendo EGB y Polimodal)



Por último, en el gráfico 6 se verá la distribución de los subsidios asignados, para cada nivel educativo alcanzado. Este diagrama de pareto permite visualizar que más del 50% de los subsidiados son quienes alcanzaron el nivel secundario.

