

# **TRABAJO PRÁCTICO PROBABILIDAD Y ESTADÍSTICA**

Accurso Agustín, Barbarroja Federico

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, dónde 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 Ordinal.
- **¿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) fue tenido en cuenta para el presente informe, representando un 54% los datos que fueron finalmente utilizados.

Para visualizar la cantidad de personas que alcanzaron cada nivel educativo, se utilizó un gráfico de barras.

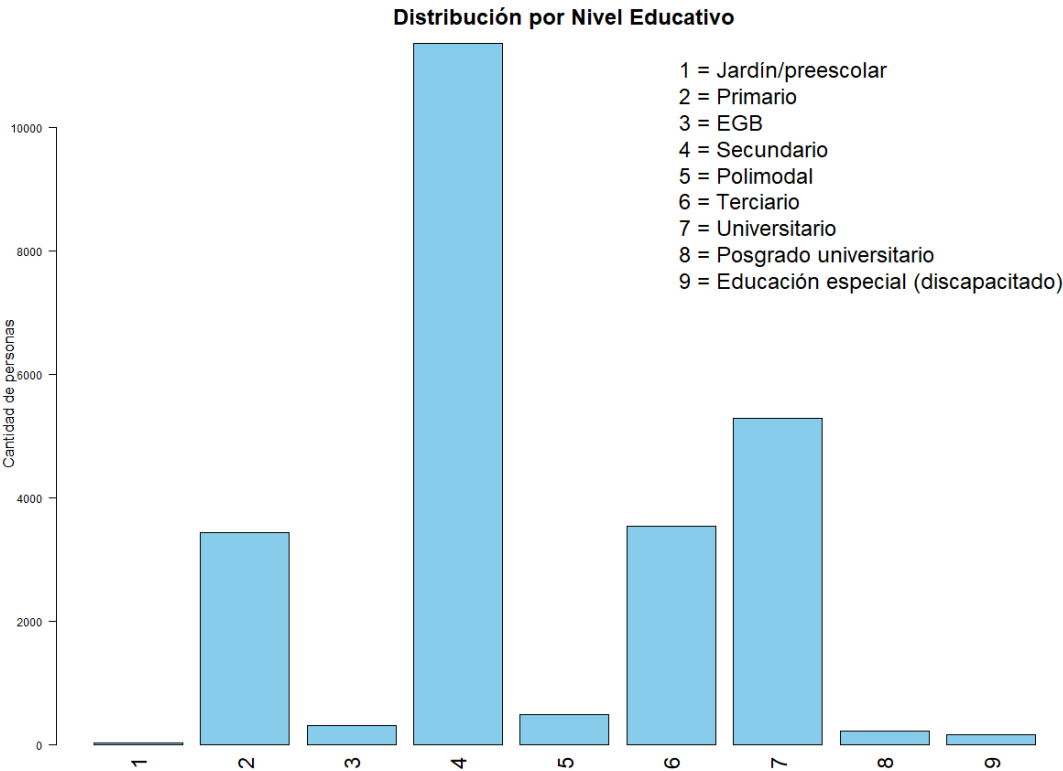


Gráfico 1.

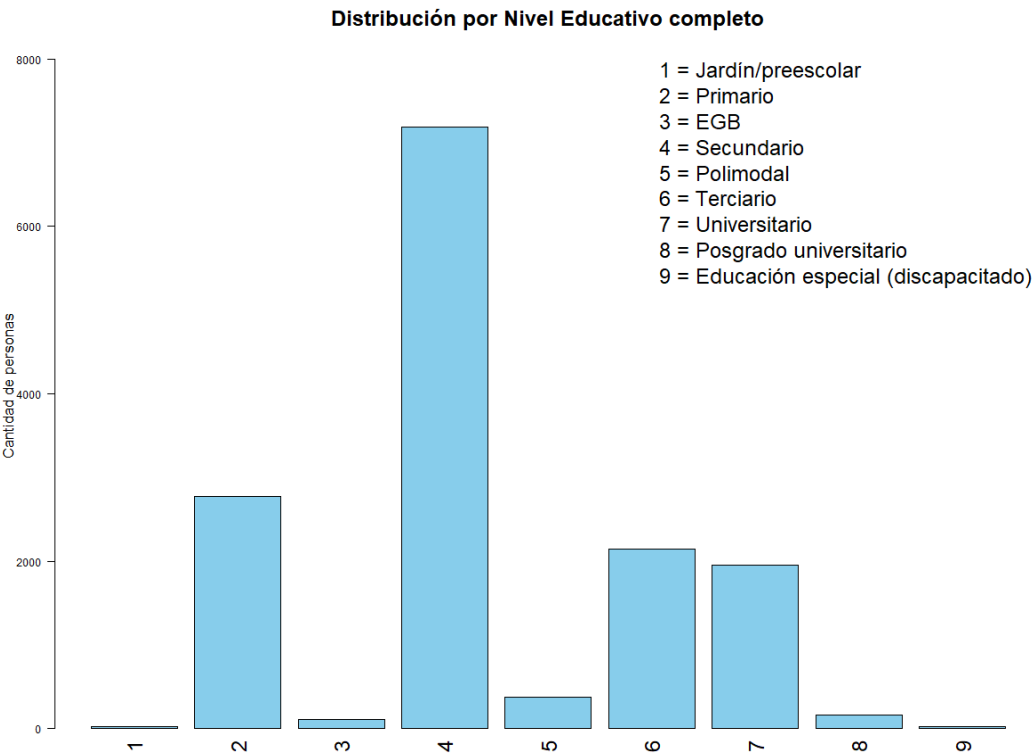
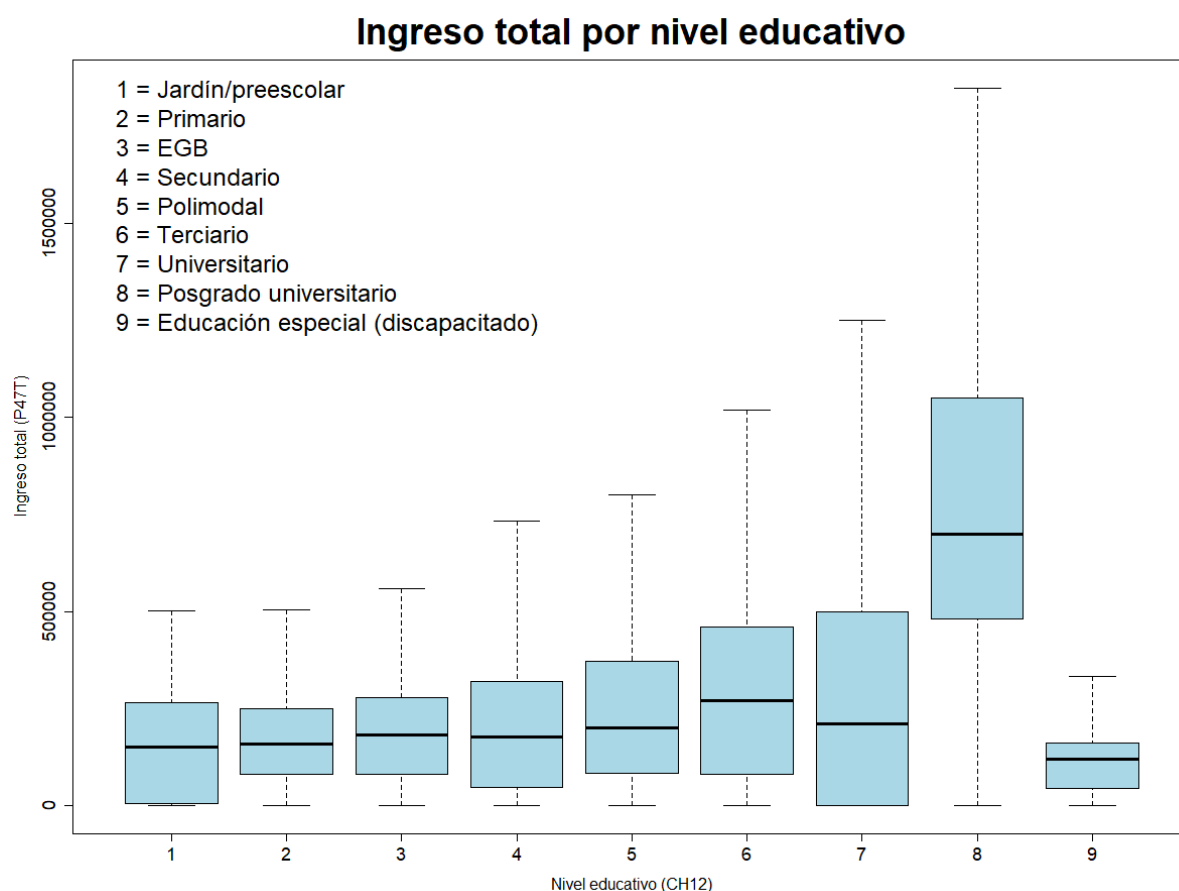


Gráfico 2.

En el gráfico 1, 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 2 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. Una hipótesis es que el nivel universitario tiene la mayor tasa de abandono por la necesidad de la gente de empezar a tener ingresos y la dificultad que conlleva el trabajo y estudio de forma simultánea.

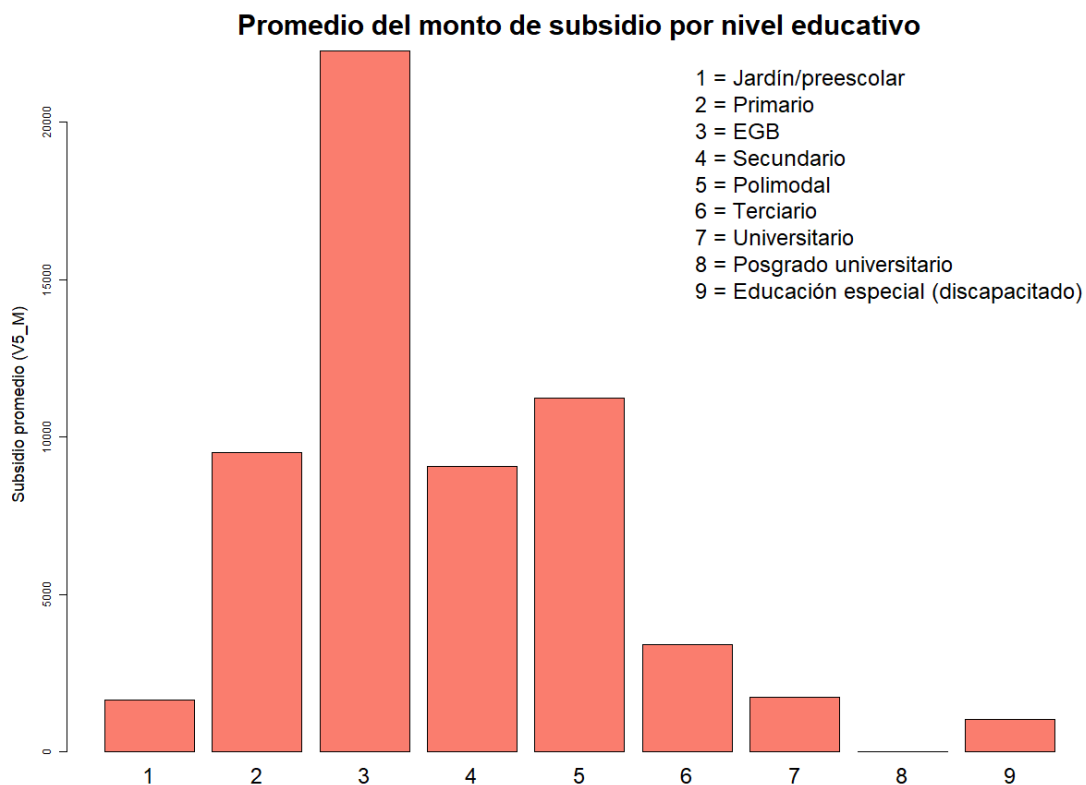
Se sigue adelante con la variable de ingreso total individual, que no será analizada por sí sola sino en relación con el nivel educativo alcanzado.



**Gráfico 3**

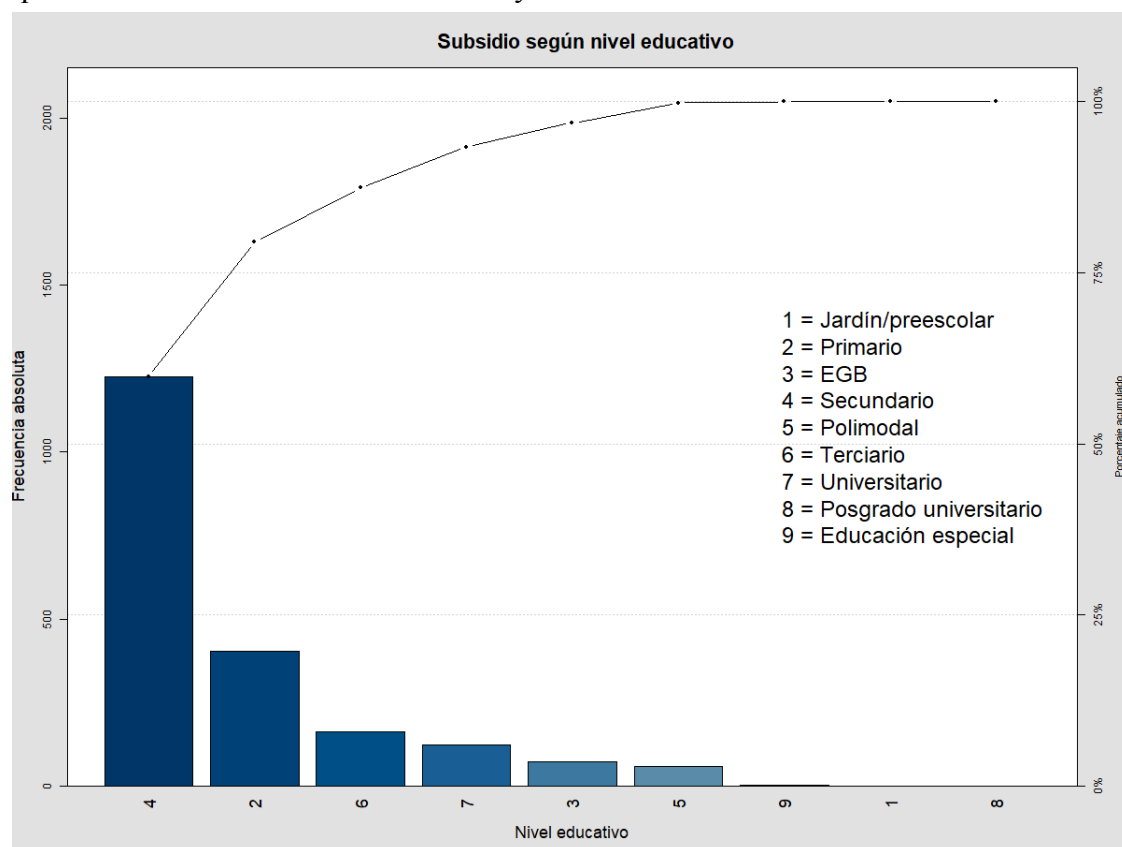
Se observa en el gráfico 3 que las personas con posgrado universitario sobresalen del resto en cuanto a sus ingresos totales. Para el presente diagrama 'boxplot' no se tuvieron en cuenta los valores atípicos porque impedían la visualización clara de los datos, los mismos se encontraban sobre todo en los niveles primario, secundario, terciario y universitario, llegando a alcanzar valores casi absurdos (\$32.600.000)

Ahora, en el gráfico 4, se analizará el monto promedio de ingresos por subsidios, también según el nivel educativo alcanzado. Vemos que el promedio del monto de ingreso por subsidios es mayor en los ciclos primarios y secundarios (incluyendo EGB y Polimodal). También vemos que, sorprendentemente, el mismo es muy bajo en la educación especial. Como es de esperarse, es casi nulo en el posgrado universitario.



**Gráfico 4.**

Por último, en el gráfico 5 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 y no continuaron sus estudios.



**Gráfico 5.**

## Parte 2

Respecto a las variables elegidas, como primer análisis se decide calcular intervalos de confianza sobre los ingresos dependiendo del nivel educativo alcanzado. Al ser varios niveles pueden surgir interesantes observaciones.

En primer lugar, se utiliza R para calcular medias de ingreso por nivel educativo y un intervalo de confianza del 95% para esa media, además de contar que tan grande es el subconjunto que se obtuvo en cada nivel. El resultado es el siguiente:

Nivel educativo	Media (\$)	IC 95%	Tamaño de subconjunto
1 – Jardín/preescolar	174.629,6	( 111.318 ; 237.941,2 )	27
2 – Primario	191.391	( 184.799,3 ; 197.982,8 )	3.445
3 – EGB	196.780,5	( 178.432,5 ; 215.128,4 )	316
4 – Secundario	229.981,4	( 224.384,1 ; 235.578,7 )	11.366
5 – Polimodal	249.378,2	( 226.910,4 ; 271.846 )	492
6 – Terciario	323.477	( 312.214,7 ; 334.739,2 )	3.546
7 – Universitario	339.513,1	( 321.955 ; 357.071,2 )	5.294
8 – Posgrado universitario	846.520,7	( 761.192,1 ; 931.849,4 )	217
9 – Educación especial	117.148	( 90.700,18 ; 143.595,8 )	157

### Observaciones:

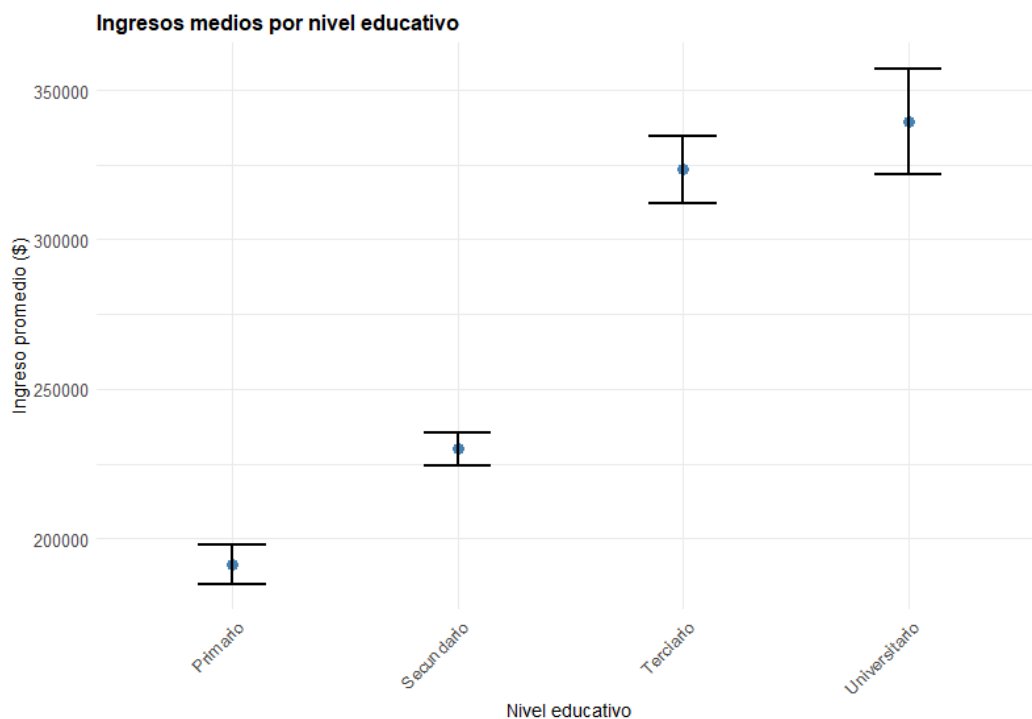
- La media tiene una tendencia creciente a medida que crece el nivel educativo.
- Quienes alcanzaron la educación especial tienen la media de ingresos más baja.
- El posgrado universitario tiene una media mucho más alta que el resto, pero teniendo en cuenta la amplitud del intervalo, podemos decir que hay una gran dispersión, debido al tamaño de subconjunto relativamente pequeño.
- Pasa algo parecido con el grupo jardín y Educación especial, tienen tamaños de muestras muy chicos, por lo que sus intervalos son muy amplios.
- Los intervalos de secundario y terciario no se superponen, por lo que vemos que el grupo terciario tiene diferencias significativas reales en comparación con el grupo secundario.

### Conclusiones sobre la media:

Para las conclusiones vamos a diferenciar las categorías, ya que consideramos que algunas de ellas no cuentan con un tamaño de subconjunto adecuado para una conclusión robusta.

Para un análisis preciso, elegimos las categorías primario, secundario, terciario y universitario: Los IC de los niveles universitario y posgrado no se superponen con los de niveles primario o secundario, lo que sugiere una diferencia estadísticamente significativa en el ingreso medio entre estos grupos.

Si bien la media del nivel universitario es mayor que la del nivel terciario, vemos que sus IC se superponen, por lo que no podemos decir que la media de los ingresos de los universitarios sea estrictamente mayor al de los universitarios. Cosa que no pasa en el nivel primario y secundario, donde la media es mayor en el grupo secundario y además sus IC no se superponen, evidenciando una media de ingreso superior por parte del nivel secundario.



**Gráfico 6.**

En segundo lugar, se calcula la proporción de personas que reciben subsidios según su nivel educativo alcanzado, con un intervalo de confianza del 95%. El subconjunto tenido en cuenta es igual al de la media de ingresos. El resultado fue el siguiente:

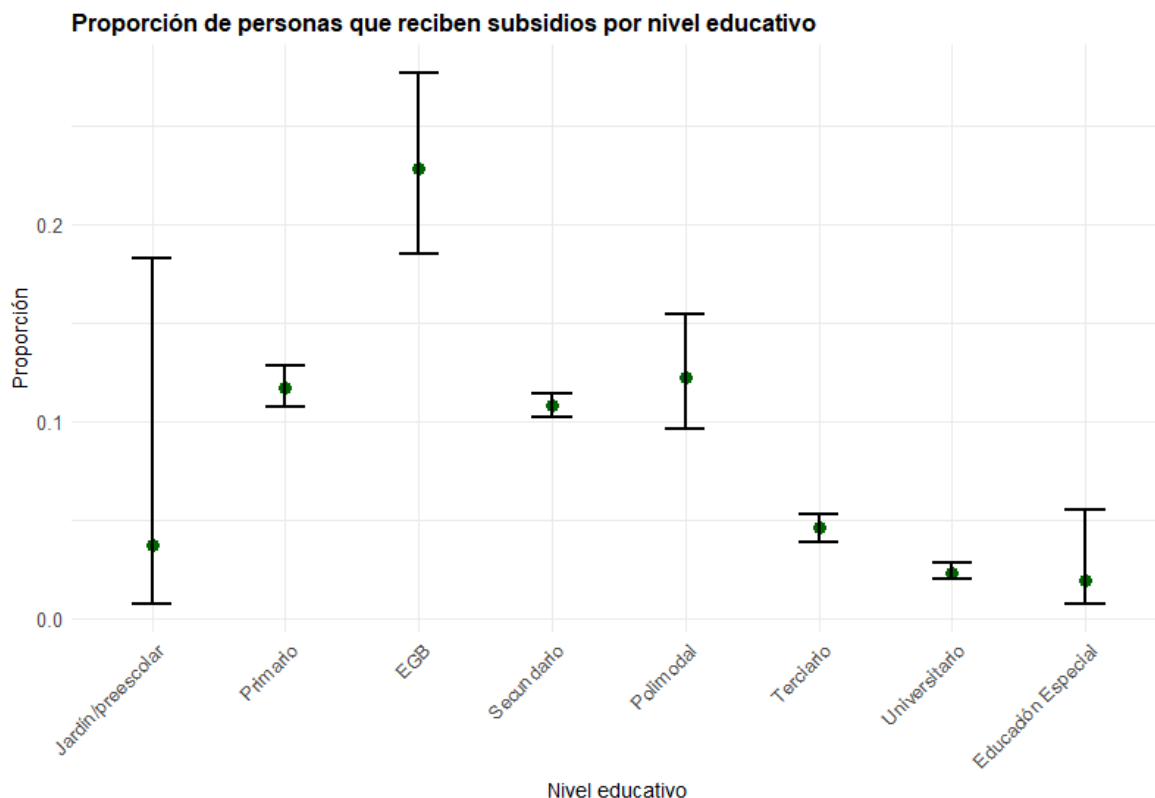
Nivel educativo	Proporción	IC 95%	Tamaño de subconjunto
1 – Jardín/preescolar	0,037	( 0,007 ; 0,183 )	27
2 – Primario	0,117	( 0,107 ; 0,128 )	3.445
3 – EGB	0,228	( 0,185 ; 0,277 )	316
4 – Secundario	0,108	( 0,102 ; 0,114 )	11.366
5 – Polimodal	0,122	( 0,096 ; 0,154 )	492
6 – Terciario	0,046	( 0,039 ; 0,053 )	3.546
7 – Universitario	0,023	(0,020 ; 0,028 )	5.294
9 – Educación especial	0,019	( 0,007 ; 0,055 )	157

Podemos observar que el nivel ‘8 – Posgrado universitario’ no figura en la tabla ya que no cumplió con el requisito mínimo, que fue tener un tamaño de subconjunto mayor a 10 y casos de éxito (personas con valor de subsidio superior a 0) mayores a 0.

Finalmente se presenta el gráfico con las proporciones y los IC obtenidos:

### **Conclusiones sobre la proporción de subsidiados:**

Como vemos en la tabla y el gráfico 7, las personas que recibieron una educación especial son en proporción los menos alcanzados por subsidios



**Gráfico 7**

### Conclusiones finales:

En resumen, el análisis permite afirmar que existe una relación positiva entre el nivel educativo alcanzado y el ingreso promedio, con diferencias que resultan significativas en varias categorías. Podríamos decir con un 95% de confianza que quienes alcanzaron el nivel secundario tienen en promedio ingresos mayores a quienes alcanzaron solo el primario. La misma diferencia sucede entre nivel secundario y terciario, pero no así para este último frente al nivel universitario.

Si bien ciertos grupos (como posgrado o educación especial) requieren un análisis más profundo debido a su tamaño, las comparaciones realizadas entre niveles educativos comunes (como primario, secundario, terciario y universitario) aportan evidencia sólida y estadísticamente respaldada.

Por otro lado, aunque no sea parte del análisis exhaustivo, nos sorprendieron 2 cosas.

1. La diferencia de ingresos promedio del grupo Posgrado universitario, que aunque tengan un IC amplio, es notablemente superior al resto.
2. Lo bajo que es el ingreso promedio de la categoría Educación especial.