

Los años de escolaridad según la pobreza multidimensional, el nivel de escolaridad de la madre y el sexo: un análisis de regresión múltiple

Pre-registro

2021-10-07

Para la elaboración de este pre-registro se ha seguido la plantilla de preinscripción de AsPredicted.org.

Gabriela Ruz, Fabian Alarcon, Martin Henriquez, Nicolás Gajardo y Gabriela Zenteno.

Resumen

En la sociedad chilena, la educación constituye un valor muy apreciado por la población, que ve en ella la única forma de progresar en la vida. Por ello, la mayoría de las personas se encuentran dispuestas a realizar grandes esfuerzos para obtener un buen nivel educacional (Oliva, 2018). Sin embargo, existen factores vinculados a desigualdades socioeconómicas y demográficas de origen que obstruyen la posibilidad de dedicar más tiempo de vida a los estudios, de obtener un buen desempeño educativo, o de prevenir la deserción o retraso. La presente investigación busca analizar en que medida las variables nivel de escolaridad alcanzado por la madre, sexo y pobreza multidimensional, influyen en los años de escolaridad de las personas, entendiendo que dichas variables proveen información sobre las circunstancias socioeconómicas que caracterizarán la trayectoria educacional. Se elaborará un modelo de regresión múltiple a partir de los datos proporcionados por la encuesta CASEN 2017.

¿Se han recopilado ya datos para este estudio?

Los datos a utilizar para esta investigación fueron extraídos de un estudio a cargo del Ministerio de Desarrollo Social, a partir de la Encuesta de Caracterización Socioeconómica CASEN 2017 que integra siete módulos; registro de residentes, educación, trabajo, ingresos, salud, redes y participación, vivienda y entorno e identidades (CASEN, 2017). El objetivo de la encuesta es conocer periódicamente la situación de los hogares y de la población en situación de pobreza y grupos definidos como prioritarios. La cobertura geográfica del estudio es a nivel nacional y corresponde a un total de 68.466 viviendas efectivas, 70.948 hogares, 216.439 personas perteneciente a hogares y 83.232 núcleos familiares. La base cuenta con 804 variables en total. Para efectos de la presente investigación, la base filtrada¹ que se utilizará cuenta con 5 variables, 1 dependiente, 2 predictores y 2 de control contando con 51.472 casos válidos (individuos).

¿Cuál es la pregunta principal o la hipótesis que se está probando en este estudio?

El presente trabajo se propone indagar en la relación existente entre la variable dependiente **años de escolaridad** de la población chilena, y factores de índole socioeconómico y sociodemográfico, a saber; **nivel de escolaridad alcanzado por la madre, sexo y pobreza**

¹ No se contabilizan los casos NA.

multidimensional. Proponemos que estas variables influyen directamente en los años de escolaridad que alcanzan las personas, así como también sus posibilidades para encontrar empleos y tener buenos salarios.

Hipótesis

H0: 1. Se plantea que los años de escolaridad tienden a ser mayores en personas que presenten situación de pobreza multidimensional. 2. Se plantea que los años de escolaridad tienden a ser mayores en personas que presenten menor nivel de escolaridad de la madre o figura materna. 3. Se plantea que los años de escolaridad tienden a ser mayores en personas se declaren de sexo femenino.

H1: Se plantea que los años de escolaridad tienden a ser mayores en personas que no presenten situación de pobreza multidimensional.

H2: Se plantea que los años de escolaridad tienden a ser mayores en personas que presenten mayor nivel de escolaridad de la madre o figura materna

H3: Se plantea que los años de escolaridad tienden a ser mayores en personas se declaren de sexo masculino.

Describa la(s) variable(s) clave(s) especificando cómo se medirán.

Se consideró la variable dependiente “Años de escolaridad” (esc), variable continua que consiste en el número de años de estudio de una persona. Luego, las variables predictoras son “Nivel de educación más alto alcanzado por la madre” (r12a), considerando 5 niveles que se encuentran indicados en la Tabla 1, recodificados a partir de la categorización CINE de la UNESCO (2011), donde

1 = Primaria incompleta o menos

2 = Primaria y secundaria baja

3 = Secundaria alta

4 = Terciaria de ciclo corto

5 = Terciaria y postgrado

También se encuentra el predictor “Situación de pobreza multidimensional” (pobreza_multi_5d), que es una medición basada en 5 dimensiones (Educación, Salud, Trabajo y Seguridad Social, Vivienda y Entorno, Redes y Cohesión Social), y que se expresa como variable dummy (0 = No es pobre; 1 = Si es pobre). En tercer lugar, se incluye la variable “sexo” dicotómica con categorías Hombre (0) y Mujer (1), y control por “edad” como variable continua.

La manera en que se irán trabajando las variables será desde el software Rstudio, ya que otorga facilidades en la codificación de las variables y gráficos, cuenta con un lenguaje sencillo de programación, así como también el de visualizar los productos que se van generando desde la base de datos, permitiendo monitorear los avances en función de los objetivos propuestos en la investigación.

Especifique exactamente qué análisis se realizará para examinar la pregunta / hipótesis principal

Con el objetivo de dar cuenta de la capacidad predictiva de las variables independientes frente a la dependiente, en primer momento se efectuará una descripción de cada una, con sus respectivos estadísticos. Luego, se procederá con un análisis bivariado que considere las correlaciones entre estas, incluyendo gráficos de caja para las variables dicotómicas. Finalmente se elaborará un modelo de regresión múltiple, que considerará de manera conjunta a las variables independientes para predecir los años de escolaridad. Se representa mediante la siguiente ecuación:

Modelo Regresión Múltiple

Años de escolaridad = $\beta_0 + \beta_1 \cdot \text{pobreza multidimensional} + \beta_2 \cdot \text{nivel de escolaridad de la madre} + \beta_3 \cdot \text{sexo} + \beta_4 \cdot \text{edad} + \varepsilon$

Donde:

- Años de escolaridad es la variable a predecir
- β_0 es el intercepto
- β_i es el coeficiente de regresión, que indica cuánto varían los años de escolaridad por cada punto que aumentan las variables independientes.
- ε el error del modelo de regresión.

¿Algún análisis secundario?

No se realizará ningún análisis secundario.

¿Cuántas observaciones se recopilarán o que determinará el tamaño de la muestra? No es necesario justificar la decisión, pero sea preciso sobre cómo se determinará exactamente el número.

El tamaño de la muestra es de 51.472 casos válidos (individuos), ya que, se excluyeron los NA de las variables de interés.

¿Algo más que le gustaría agregar? (por ejemplo, exclusiones de datos, variables recopiladas con fines exploratorios, análisis inusuales previstos)

Se utilizaron datos recopilados por la encuesta CASEN 2017, pero se filtraron ciertas variables de interés para fines de esta investigación. La base original cuenta con 804 variables en total. Para efectos de la presente investigación, la base filtrada que se utilizará cuenta con 5 variables, 1 dependiente, 2 predictores y 2 de control contando con 51.472 casos válidos (individuos).

Respecto al nivel de apertura y documentación de los datos, tomando como referencia los estándares FAIR, nuestra base de datos cumple con todos los parámetros establecidos 1) Respecto a si se puede encontrar, CASEN (2017) se encuentra disponible en la página del ministerio del desarrollo social; 2) Respecto a la accesibilidad, es una base de datos de uso público y gratuito, sin necesidad de acreditaciones para acceder a los datos; 3) Respecto a interoperabilidad, la base de datos está codificada de tal manera que pueda ser comparada con otras de años anteriores, esto gracias a la consistencia en los libros de códigos. 4) Por último,

sobre la reusabilidad, las variables se encuentran específicamente descritas en los anexos.

Finalmente, sobre las indicaciones de LISA para apertura de datos, en la página web de Observatorio social de Previsión social que se encuentra en referencias, la base de datos utilizada contiene la información recopilada en distintos formatos, como por ejemplo, se encuentra disponible en *.sav* (SPSS) y *.dta* (STATA). Como también, en caso de estar trabajando con una encuesta, es el archivo que contiene las preguntas que se le realizaron a los participantes del estudio. Es un documento de texto que está en formato *.pdf*. En cuanto al cuestionario, el documento donde aparece información de las variables, valores y etiquetas, correspondientes a la base de datos, se puede ver indicada la asociación entre las preguntas del cuestionario y las variables de la base de datos está disponible en formato *.pdf*. Sobre la ficha técnica o el “informe metodológico”, permite conocer características del diseño metodológico propias del instrumento de análisis., este se encuentra en formato *.pdf* (LISA-COES, 2021).

Referencias

Encuesta de caracterización socioeconómica (2017). Observatorio social. Previsión social: síntesis de resultados. Santiago de Chile: Ministerio de Desarrollo Social y Familia. Recuperado de:
<http://observatorio.ministeriodesarrollosocial.gob.cl/encuesta-casen-2017>

GO FAIR. (2016). FAIR principles. Recuperado en <https://www.go-fair.org/fair-principles/>

Instituto de Estadística UNESCO. (2011). *Revisión de la clasificación internacional normalizada de la educación (CINE)*. Paris. Recuperado en <http://uis.unesco.org/sites/default/files/documents/isced-2011-sp.pdf>

LISA-COES.(2020). *Datos abiertos, Guía para la apertura de datos*. Recuperado en <https://lisa-coes.netlify.app/02componentes/>

Oliva, H. (15 de enero del 2018) El 88% de los chilenos termina la educación media. *Centro de la prensa*. Recuperado en:
<https://gobierno.udd.cl/cpp/noticias/2018/01/15/el-88-de-los-chilenos-termina-la-educacion-media-la-tercera/>