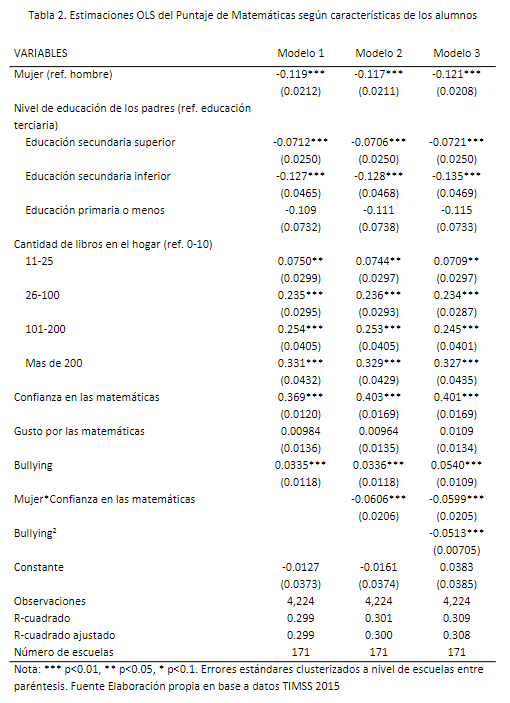
***ENTREGA 3 DISEÑO DE LA INVESTIGACIÓN SOCIAL***

Integrantes: Benjamín Carreño Canales y Ramón Jara López

**Parte I**

Un grupo de jóvenes investigadores indagó en los factores sociales y psicosociales que inciden en el desempeño en matemáticas entre alumnos de 8° Básico. Para lograr este objetivo, utilizan datos del Estudio Internacional de Tendencias en Matemática y Ciencias (TIMSS) del año 2015. La Tabla 2 presenta tres modelos de regresión lineal OLS.



Las variables corresponden a 1) Confianza: puntaje estandarizado que indica qué tan fácil perciben los alumnos que les resultará desempeñarse bien en matemáticas; 2) Gusto: puntaje estandarizado que indica qué tanto le gusta al alumno o alumna estudiar matemática; 3) Bullying: puntaje estandarizado que indica el nivel de acoso escolar del cual el alumno es víctima.

**1. Describa e interprete los coeficientes de la variable Género y Nivel de Educación de los Padres en el Modelo 1.**

Ser mujer implica una disminución en el puntaje de matemáticas en 0.119 unidades de medición con respecto al puntaje de los hombres, controlando por el resto de las variables. Además, el coeficiente es estadísticamente significativo con un valor p menor a 0.01.

Con respecto al nivel de educación de los padres, el puntaje de los estudiantes en matemáticas que tienen padres con educación secundaria superior disminuye en 0.0712 unidades en comparación con los estudiantes con padres que tienen educación terciaria, dejando el resto de las variables constantes. Por otro lado, los estudiantes que tienen padres con educación secundaria inferior disminuyen su puntaje de matemáticas en 0.127 unidades en comparación con los estudiantes que tienen padres con educación terciaria, ceteris paribus. Ambos coeficientes son significativos con un valor p menor a 0.01. Finalmente, para los estudiantes que tienen padres con educación primaria o menor un aumento de una unidad en la escala de medición disminuye sus resultados de matemáticas en promedio 0.109 puntos, a diferencia de los resultados de estudiantes con padres que tienen educación terciaria. A pesar de esto, este último valor no es estadísticamente significativo.

**2. Describa e interprete los coeficientes de las variables Escala de Confianza y Escala de Gusto por las matemáticas en el Modelo 1.**

Una desviación estándar más de confianza en las matemáticas implica un aumento de 0.396 unidades en el puntaje de matemática (coeficiente estadísticamente significativo con un valor p menor a 0.01). Acerca del gusto de las matemáticas, una desviación estándar más implica un aumento de 0.00984 en el puntaje de matemáticas. Eso sí, el valor no es estadísticamente significativo (p>0.05).

**3. Uno de los investigadores propuso incluir un efecto de interacción entre género y confianza en las matemáticas en el Modelo 2. ¿Qué está testeando con el término interacción?**

El término de interacción introducido por uno de los investigadores busca testear cómo la relación entre la confianza en las matemáticas y el puntaje en matemáticas varía para hombres y para mujeres (efecto moderador del género). En otras palabras, si es que existe un efecto significativo de la confianza en las matemáticas, condicionado por el género, en el puntaje en matemáticas.

**4. Calcule el efecto interacción para hombres y mujeres. Interprete sus resultados.**

El efecto interacción para hombres se calcula considerando el modelo interacción (modelo 2) en aquellos casos en que la variable toma valores 0 (dado que se asume categoría de referencia, donde categoría mujer es 1). Con ello, la variable interacción propiamente tal es 0 (0\*confianza en matemáticas) y se considera, en vez, el coeficiente de la variable confianza en matemáticas solamente (b1=0,403). Así bien, esto se interpreta como que el efecto adicional de confianza en matemáticas en los hombres significa un incremento de 0,403 en el puntaje de matemáticas.

Por otro lado, el efecto interacción para las mujeres se advierte al considerar el coeficiente de interacción mujer y confianza en matemáticas (1\*confianza), el que da un valor de b2=–0,0606. Esto se interpreta como que el efecto adicional de la confianza en matemáticas en mujeres significa una disminución de 0,0606 en el puntaje de matemáticas. Ambos coeficientes, para hombres y mujeres, se consideran dado que son estadísticamente significativos al 99% (valor-p menor a 0,01).

**5. Finalmente, uno de los investigadores plantea que quizás el efecto medido por la escala de bullying no sea lineal y que deberían incluir un efecto cuadrático. A partir de lo presentado en el Modelo 3, describa el efecto de la variable bullying.**

La variable bullying para el caso señalado (modelo 3) presenta dos coeficientes importantes a considerar. En un primer momento, se caracteriza el efecto lineal del bullying con el coeficiente simple reportado de 0.0540. En este caso, como la dirección es positiva y la magnitud es estadísticamente significativa (p>0.1), se interpreta que, linealmente, un incremento de una desviación estándar en el puntaje estandarizado de percepción de acoso contribuye en un aumento del puntaje de matemáticas en 0,054 puntos.

Sin embargo, al considerar ahora un efecto cuadrático en la variable bullying se advierte que el efecto de este no estaba totalmente considerado en la adecuación lineal. El coeficiente cuadrático del bullying es de –0,0513, lo que advierte de una dirección opuesta a la lineal y por tanto la interpretación varía (el coeficiente sigue siendo estadísticamente significativo). En este sentido, mientras más incremento exista en la medición de la variable bullying (es decir, en aquellos estudiantes que perciben mayor nivel de acoso) la direccionalidad compromete un claro factor negativo, lo que induce a interpretar que los puntajes de matemática serán menores en aquellos que perciben altos niveles de acoso y victimización. Si en un principio, percepciones bajas de bullying incrementan el puntaje en matemática, cuando esta percepción (la medición de la variable) aumenta, el efecto en los puntajes hace que estos disminuyan.

Por ejemplo, si asumimos un valor de bullying igual a 2 (alto en un puntaje estandarizado), se tiene, ceteris paribus, que el efecto del bullying estará dado por (0,054)\*2 + (-0,513)\*2\*2 = 0,108 – 3.026 = -2,918. Es decir, una alta percepción de acoso (bullying=2) significa un efecto negativo y grande en magnitud (-2,9) sobre el puntaje de matemáticas obtenido, controlando por el resto de las variables.

**Parte II**

**1. Considere el último modelo multivariado de regresión propuesto.**

En el análisis de la investigación en curso, la variable dependiente es el ingreso total per cápita del hogar (variable continua). La variable independiente principal es la escolaridad (medida en años), además de contar con las variables de control i) pertenencia a un pueblo originario, ii) género y iii) situación de convivencia con pareja (recodificación del estado civil). Esta última se categorizó de forma binaria, donde 0 refiere a las personas que están en pareja (incluido casados y firmantes del AUC) y 1 a quienes están solos (solteros, separados, viudos).

La pregunta de investigación se formula de la siguiente manera: ¿Cómo influye la escolaridad, considerando la pertenencia a un pueblo originario, el género y la situación de convivencia, en el ingreso total per cápita del hogar en Chile? En este contexto, se busca analizar el impacto de la escolaridad en el ingreso per cápita del hogar controlando por género, pueblo originario y estado de convivencia.

Esta formulación de la pregunta de investigación asegura que la variable dependiente, el ingreso total per cápita del hogar, esté claramente identificada, y las variables independientes y de control estén correctamente definidas para abordar la investigación propuesta. La hipótesis central, que se comprueba significativamente con el primer modelo, apela a que existe una relación positiva entre los años de escolaridad de quien aporta los mayores ingresos y los ingresos per cápita dentro del hogar. La inclusión de variables de control (género, pertenencia a pueblo originario y estado de convivencia) debieran especificar la relación central.

Tabla 0. Estadísticos descriptivos de variables a utilizar

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Variables Continuas | Media | Mediana | Desviación estándar | Mínimo | Máximo | Percentil 25 | Percentil 75 |
| Ingreso total per cápita (ypc) | 471122.4 | 349778 | 521447.8 | 68112 | 2400000 | 231158 | 537365 |
| Años de escolaridad (esc) | 11.16742 | 12 | 4.269671 | 0 | 19 | 8 | 14 |
|  |  |  |  |  |  |  |  |
| Variables Categóricas | Categoría | Frecuencia | Porcentaje |  |  |  |  |
| Sexo (genero\_binario) | Masculino | 95656 | 47.30% |  |  |  |  |
|  | Femenino | 106575 | 52.70% |  |  |  |  |
| Pueblos Originarios (p\_originario) | Sí | 28824 | 14.25% |  |  |  |  |
|  | No | 173407 | 85.75% |  |  |  |  |
| Estado Civil o Conyugal (econvivencia\_binario) | Con pareja | 83209 | 49.19% |  |  |  |  |
|  | Sin pareja | 85942 | 50.81% |  |  |  |  |

**2. Proponga un segundo modelo de regresión (Modelo 2), el que debe incluir un efecto moderador.**

**A) Defina la hipótesis que está testeando con la inclusión del término interacción en el Modelo 2. Identifique cual es la variable dependiente (Y), moderadora (X2) e independiente principal (X1) en este análisis.**

Respecto al modelo 2 (ver Tabla 1, modelo 2), se considera el efecto interacción existente entre los años de escolaridad (continua) y el género (hombre/mujer). Como tal, se define que la hipótesis a testear refiere a que el efecto de los años de escolaridad en el ingreso per cápita del hogar está condicionado por el género. Se busca testear si los años de escolaridad difieren en el efecto de ingresos si se consideran años de escolaridad de hombres y años de escolaridad de mujeres; claramente suponiendo que para este último grupo (mujeres y años escolares) el efecto será menor en los ingresos del hogar, en comparación con los hombres.

Así es que se identifica, para tal modelo, que la variable dependiente (Y) es el ingreso per cápita, la variable moderadora es la interacción entre años de escolaridad y género (X2), y la variable independiente principal (X1) refiere a los años de escolaridad.

**B) Estime el efecto interacción entre las dos variables de interés. Una de las variables incluidas en esta interacción debe ser continua.**

La estimación del efecto interacción entre los años de escolaridad (continua) y el género (hombre/mujer) es de –11.642 (ver Tabla 1, modelo 2).

**C) Interprete los resultados relativos al término interacción. ¿Es distinto el “efecto” de la variable independiente (X1) sobre la variable dependiente (Y) considerando la variable moderadora (X2)?**

El coeficiente reportado de –11.642 indica que para las mujeres el efecto adicional por cada año de escolaridad es de 11.642 pesos menos que para los hombres, ceteris paribus. Es decir, por cada año de escolaridad, las mujeres reciben menor cantidad de dinero en comparación a los hombres (en específico, 11.642 pesos por año escolar). Se trata de un coeficiente estadísticamente significativo dado que el valor-p reportado es menor a 0,01, y, por tanto, se confirma la hipótesis.

Además, se puede notar que el efecto de la variable independiente (X1) años de escolaridad es mayor al considerar la variable moderadora entre años de escolaridad y género. En concreto, la magnitud del coeficiente de los años de escolaridad aumentó de 33.904 unidades en el modelo 1 a 40.336 unidades en el modelo 2, que se traduce en que cuando no existe efecto interacción (género=0=hombre) la variable independiente principal está caracterizando el efecto específico de los hombres en el ingreso per cápita (que tiene sentido que sea superior por la direccionalidad del efecto interacción).

**D) Genere un gráfico que muestre los valores predichos para el termino interacción, y refiérase sustantivamente a estos resultados. El grafico debe ser editado.**

A partir del gráfico de los valores predichos para el término interacción (gráfico 1) se desprende que efectivamente, a un 95% de confianza, existen diferencias sustantivas entre hombres y mujeres respecto al efecto en los ingresos de los años de escolaridad alcanzados por cada uno. Diferencia que comienza a notar su significación a partir de los 10-12 años de escolaridad, lo que permite intuir que, con anterioridad, no existe efecto condicionador del ser hombre o mujer (lo que se explica en parte en que los años de escolaridad en los primeros intervalos –escolaridad básica y media- no tiene mayor efecto en los valores predichos del ingreso per cápita del hogar).

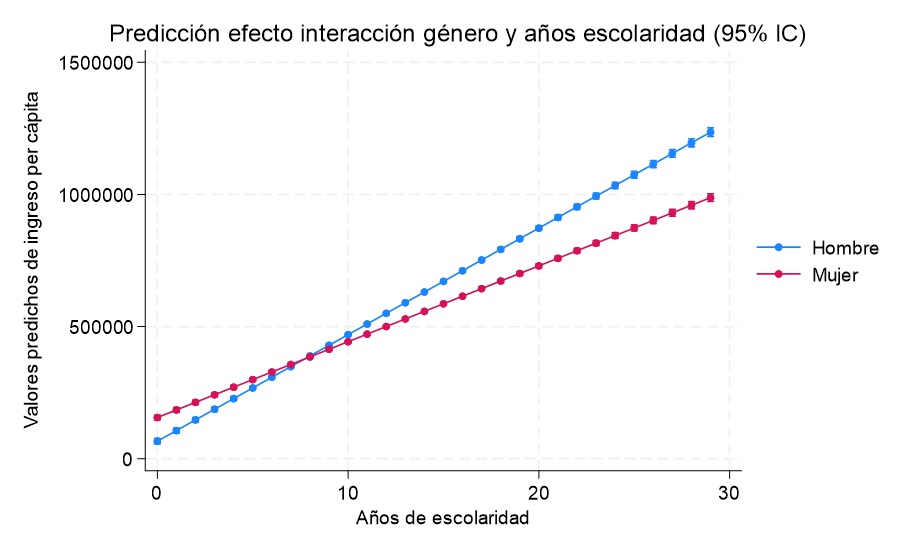


Gráfico 1. Valores predichos de efecto interacción género-escolaridad

**E) Compare el ajuste del Modelo 1 respecto del Modelo 2.**

El ajuste en este caso ha aumentado levemente. El modelo 1 reporta un R cuadrado de 0.077, lo que significa que el modelo explica un 7.7% de la variabilidad de la variable dependiente. Por otro lado, el modelo 2 reporta un R cuadrado de 0.079, es decir, el modelo 2 explica un 7.9% de la variabilidad del ingreso per cápita. En tal caso es que se considera un mejor ajuste en el modelo 2, pero con una diferencia pequeña.

**3. Proponga un tercer modelo de regresión (Modelo 3), el que debe incluir un nuevo efecto moderador.**

**A) Estime el efecto interacción entre las dos variables de interés.**

La estimación del efecto interacción entre la variable género y la variable pertenencia a pueblos originarios (ambas categóricas) es de 4,853 (ver Tabla 1, modelo 3).

**B) Defina la hipótesis que está testeando con la inclusión del término interacción en el Modelo 3. Identifique cual es la variable dependiente (Y), moderadora (X2) e independiente principal (X1) en este análisis.**

La hipótesis considerada en este modelo intenta testear si es que existe un efecto interacción a partir del género y la pertenencia a pueblos originarios. En específico, se testea la idea de que la tendencia general de los ingresos se ve afectada por la pertenencia a pueblos originarios, a su vez condicionada por el género. En otras palabras, si existe variación en el efecto en los ingresos a partir de la pertenencia a pueblos originarios, diferenciando por hombres y mujeres pertenecientes y no pertenecientes.

La variable dependiente (Y) sigue siendo el ingreso per cápita del hogar, la variable independiente principal (X1) es años de escolaridad y, para este caso, la variable interacción (X2) está dada por la pertenencia a pueblos originarios y el género.

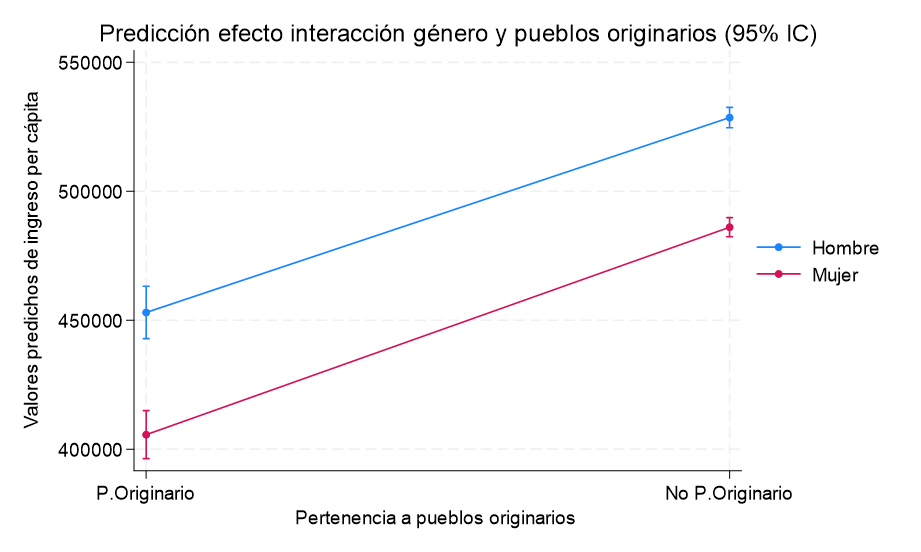
**C) Interprete los resultados relativos al término interacción.**

El coeficiente que reporta el modelo es de 4.853, lo que implica que para las mujeres que no pertenecen a pueblos originarios, el efecto adicional sobre los ingresos per cápita del hogar es de 4.853 pesos más para las mujeres que sí forman parte de los pueblos originarios, ceteris paribus. El coeficiente no es estadísticamente significativo al considerar su valor-p mayor a 0,01, y por tanto se rechaza la hipótesis planteada anteriormente.

Aun así, cabe destacar que el efecto de los hombres que pertenecen a pueblos originarios (genero=0; pertenencia=1) es, en promedio, 75.587 pesos menos en los ingresos per cápita del hogar en comparación a los hombres que no pertenecen a pueblos originarios, controlando por el resto de las variables.

**D) Genere un gráfico que muestre los valores predichos para el termino interacción, y refiérase sustantivamente a estos resultados. El grafico debe ser editado.**

A partir del gráfico que reporta los valores predichos para la interacción entre el género y la pertenencia a pueblos originarios (ambas variables categóricas) se observa que, considerando un 95 % de confianza, no existen diferencias significativas en el efecto en los ingresos considerando la inclusión de pueblos originarios condicionada por género. Si bien las diferencias categóricas son claras gráficamente, estas no se comportan siguiendo un patrón de interacción (líneas convergentes en un punto), y por tanto no es significativo el efecto interacción. En el fondo, se evidencia que el efecto en los ingresos de la pertenencia a pueblos originarios es significativo, pero independiente de una moderación del género.



**E) Compare el ajuste del Modelo 1 respecto del Modelo 3.**

Como se mencionó anteriormente el ajuste del modelo 1 reporta un R cuadrado de 0.077. En el caso del modelo 3 hay una disminución con respecto al modelo 2, porque presenta el mismo R cuadrado del modelo 1. Por lo tanto, la interpretación sería que el modelo 3 solo explica en un 7,7% la variabilidad de la variable dependiente (al igual que el modelo 1), levemente inferior al ajuste del 7,9% del modelo 2. Lo que tiene sentido al haber incluido un efecto interacción sin mayor significancia.

**4. Proponga un cuarto modelo. Llame a este Modelo 4. Incluya en este modelo una nueva transformación aritmética (polinomio o transformación logarítmica).**

**A) Defina la hipótesis que está testeando con la inclusión de esta transformación. Porque es relevante incluir esta transformación.**

La hipótesis a testear apunta a que el comportamiento de la variable años de educación, en relación con el ingreso per cápita, no es necesariamente lineal. Lo que significa, para el caso cuadrático en cuestión, que una mejor manera de evidenciar el efecto de la escolaridad en los ingresos es en la inclusión de un efecto cuadrático de los años de escolaridad. En síntesis, se testea que la incorporación de la variable dependiente principal escolaridad no obedece a un efecto lineal, sino que a un efecto cuadrático (cambia de dirección en un momento determinado y su tendencia es exponencial).

La relevancia de incorporar un efecto cuadrático de los años de escolaridad está dada por la dificultad de mantener el supuesto de que la relación lineal entre años de escolaridad e ingresos sea tal en todo intervalo de años educativos y sea siempre la misma. Con ello, se posibilita que la relación pueda ser diferencial dependiendo de los años educacionales que se consideran (el impacto en los ingresos no es igual en los primeros años de educación -media, por ejemplo- que en un año educacional adicional al considerar la educación superior –un posgrado, por ejemplo).

**B) Interprete los resultados relativos al término incluido.**

El coeficiente que indica el modelo es 5.283 para la edad al cuadrado, esto indica una tendencia curva con una forma convexa (forma de U) entre la variable independiente y dependiente. Al considerar, además, el coeficiente lineal de los años de escolaridad (-72,178) se nota que en los primeros años de escolaridad el efecto en los ingresos de un año más de escolaridad disminuye hasta que en un punto particular (a los 13 años de escolaridad, cuando b1\*esc < b2\*esc^2) el efecto cambia de dirección y un año de escolaridad más incrementa exponencialmente el efecto en los ingresos per cápita por hogar. El coeficiente reporta un valor estadísticamente significativo a un 99% de confianza y por tanto se confirma la hipótesis previa.

**C) Compare el ajuste del Modelo 1 respecto del Modelo 4.**

El modelo 4 indica una mejoría sustancial con respecto a los modelos anteriores. El R cuadrado reportado es de 0.136, lo que significa que este modelo explica en un 13,6% la variabilidad del ingreso per cápita.

**D) Desarrolle análisis de comparación de modelos. Presente los modelos en una tabla editada. Indique cual es el mejor modelo de los modelos presentados. Justifique su decisión.**

A partir de los modelos reportados en la Tabla 1 es claro que, tanto por ajuste como por justificación de la incorporación de variables y modelamiento, el modelo 4 es el mejor modelo por considerar. Por un lado, el modelo 4 refiere al mejor ajuste comparado con los otros 3. Pasa a tener un R cuadrado reportado de 0,136 (lo que está muy por sobre los R cuadrado reportados de los otros tres modelos). Por otro lado, la consideración de un efecto cuadrático, en este caso dado por los años de escolaridad al cuadrado, significa un mejor acercamiento de la tendencia general que asocia años de escolaridad e ingresos. En ese sentido, se argumenta a favor de la intuición de que en los primeros años la escolaridad no parece ser un predictor de ingresos, incluso marcando una tendencia negativa (lo que podría explicarse por el costo de oportunidad que implica la formación educativa básica-media). El efecto solo se revierte, pasando de un efecto negativo a uno positivo, con el paso de los años (a partir de los 13 educativos), considerando ahora el efecto exponencial que tiene cada año extra de escolaridad (universitario, posgrado, etc.) en los ingresos per cápita del hogar.

Tabla 1. Modelos que relacionan los ingresos y los años de escolaridad

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
| VARIABLES | Modelo 1 | Modelo 2 | Modelo 3 | Modelo 4 |
|  |  |  |  |  |
| Años de escolaridad | 33,904\*\*\* | 40,336\*\*\* | 33,905\*\*\* | -72,162\*\*\* |
|  | (301.2) | (449.2) | (301.2) | (1,039) |
| Género (Cat. Ref.: Hombre) | -40,379\*\*\* | 89,769\*\*\* | -44,582\*\*\* | -46,342\*\*\* |
|  | (2,578) | (7,226) | (7,034) | (2,495) |
| Pertenencia a pueblos originarios (Cat. Ref.: Pertenecen) | 78,217\*\*\* | 77,910\*\*\* | 75,587\*\*\* | 67,133\*\*\* |
|  | (3,774) | (3,770) | (5,569) | (3,653) |
| Estado Civil (Cat. Ref.: Con pareja) | -14,479\*\*\* | -15,618\*\*\* | -14,497\*\*\* | -15,787\*\*\* |
|  | (2,574) | (2,572) | (2,574) | (2,490) |
| Género\*Años de escolaridad |  | -11,642\*\*\* |  |  |
|  |  | (603.9) |  |  |
| Género\*Pueblos Originarios |  |  | 4,853 |  |
|  |  |  | (7,558) |  |
| Años de Escolaridad^2 |  |  |  | 5,283\*\*\* |
|  |  |  |  | (49.69) |
| Constante | 77,818\*\*\* | 6,598 | 80,104\*\*\* | 520,654\*\*\* |
|  | (5,086) | (6,281) | (6,208) | (6,446) |
|  |  |  |  |  |
| N | 165,132 | 165,132 | 165,132 | 165,132 |
| R-cuadrado | 0.077 | 0.079 | 0.077 | 0.136 |
| Error estándar entre paréntesis | |  |  |  |
| \*\*\* p<0.01, \*\* p<0.05, \* p<0.1 | |  |  |  |