# Universidad Nacional Autónoma de México Facultad de Economía

**Centro de Educación Continua y Vinculación**

# Diplomado en Econometría

**Módulo IV.** Modelos de datos de panel

# Proyecto Final

Pedro Huet López

**I.- Introducción.**

1. *Propuesta de análisis.*

Con el fin de hacer un análisis robusto en materia de modelos probabilísticos y bases de datos panel, en este proyecto se busca estudiar si las características específicas relacionadas con el rendimiento de las escuelas afectan la probabilidad de que los estudiantes abandonen sus estudios.

Mi interés de llevar a cabo este análisis proviene de mi experiencia personal: al estar estudiando algunos temas escolares en mi entidad (Estado de México), observé que había innumerables familias que sacaban a sus hijos de las escuelas. Al haberme puesto a pensar sobre el tema, me quedó el interés de saber si, más allá de la situación socioeconómica de las familias, la decisión de sacar a los hijos de la escuela podría seguir la lógica de una función de productividad: a mejor rendimiento y facilidades de una escuela (eficacia), hay menores incentivos de que una familia saque a sus hijos (bajas, o pérdidas).

Las implicaciones de este estudio son relevantes, ya que, si se llegara a ofrecer evidencia de que la deserción escolar es afectada por las características del valor que ofrecen las escuelas, podría indicar que una estrategia viable para reducir la deserción escolar es que las autoridades y comunidades inviertan y apoyen a las instituciones educativas. Por el contrario, si el estudio indica que hay poca relación entre la deserción escolar y las variables pertinentes a las escuelas, se podría concluir que el gasto en educación podría no ser tan vital para reducir la deserción escolar, comparado al gasto dirigido a mejorar la situación económica de la población.

1. *Base de datos y variables utilizadas.*

Para el estudio, se utiliza la base de datos de Panel de Escuelas Brasileñas, del Banco Interamericano de Desarrollo (Huberts y Machado, 2017). La base de datos contiene datos de más de 4 millones de escuelas de educación básica (preescolar a secundaria) públicas y privadas de 27 entidades de Brasil de 1996 a 2016 (20 años), en las cuales se pudieron recabar los datos. Por el inmenso número de observaciones y que no hubo un sesgo aparente en su selección, es probable que la distribución de la muestra sirva para reflejar una imagen confiable de la distribución de las características de las escuelas de ese país.

Para el estudio, se emplea 1 variable dependiente y 4 variables regresoras, todas provenientes de la base de datos anteriormente mencionada. En el caso de la variable dependiente, se selecciona el estado de deserción de la escuela: si la escuela sufre de alta deserción escolar o si no enfrenta esta situación (variable dicotómica). Para ello, se lleva a cabo una transformación de esta variable, originalmente de tipo continua, en la base de datos. La razón principal por la cual se elige transformar esta variable es porque, al estar originalmente codificada de forma continua (con diferentes valores por hasta 4 decimales), en este estado la variable ofrece una gama demasiado extensa de valores para poder operacionalizar su efecto. Al convertirla en dicotómica, se toma el umbral de una deserción escolar de 5% o más como una medición de deserción escolar amplia. Esta medición se eligió porque, varios especialistas sobre la educación en ese país (Saldaña, 2021) han señalado que una deserción escolar de alrededor de 5% se considera alta, ya que supera la cifra de deserción escolar nacional más alta que ha tenido Brasil en memoria reciente (mayor al 4.8% de 2006).

1. *Teoría e hipótesis.*

En el caso de las variables regresoras, se eligieron **a)** el número de profesores de educación básica que tiene la escuela (variable discreta), **b)** el promedio del número de años de escolaridad de los profesores (variable discreta), **c)** número de alumnos por salón o clase (variable discreta) y **d)** el número de computadoras que tiene la escuela (variable discreta).

En el caso del **número** y **años de escolaridad** **de los profesores** educación básica, numerosos estudios (Furger, 2008) han demostrado que, al recibir los alumnos una instrucción mejor y más personalizada de parte de los profesores, los familiares tienen mayores incentivos de mantener a sus hijos en las escuelas (tanto porque incrementa su percepción de que el costo de oportunidad de que el niño estudie está valiendo la pena, como por una mayor confianza de poderse acercar al profesor y pedir apoyo para que sus hijos puedan seguir estudiando). En el caso del **número de alumnos por salón**, numerosos especialistas (National Council of Teachers of English, 2014) han concluido que, a un mayor número de alumnos en el salón de clases el aprendizaje y rendimiento de los estudiantes disminuye a partir del aumento de distractores. Al mostrar un menor rendimiento y aprendizaje, el alumno y sus familiares podrían estar menos interesados de que su hijo concluya sus estudios. Finalmente, el **número de computadoras** sirve como un buen indicador de los recursos que destinan las escuelas para los niños: más computadoras tenga una escuela, se piensa que los alumnos tienen acceso a más recursos tecnológicos. Por ello, se esperaría que las familias y alumnos aprecien tanto estos bienes, que sus incentivos de sacar a los niños de la escuela disminuyan (Lynch, 2017).

En resumen, se esperan los siguientes resultados:

*H1: A mayor número de profesores tenga una escuela, se reducirá la deserción escolar.*

*H2: A mayor escolaridad tengan los profesores, se reducirá la deserción escolar.*

*H3: A mayor número de alumnos por salón, aumentará la deserción escolar.*

*H4: A mayor número de computadoras tenga una escuela, se reducirá la deserción escolar.*

**II.- Número de observaciones codificación precisa de las variables.**

Eliminando las observaciones sin información y balanceando el panel para analizar las mismas escuelas a través del tiempo sin brechas, se obtienen los datos de 22 escuelas en 12 años diferentes (2003 – 2014), resultando en un total de 264 observaciones. La codificación de las variables es la siguiente:

Variable dependiente:

1.- Escuela que presenta una deserción escolar alta: variable dicotómica (1, 0).

Variables independientes:

1.- Número de profesores por escuela: variable discreta (0, 100).

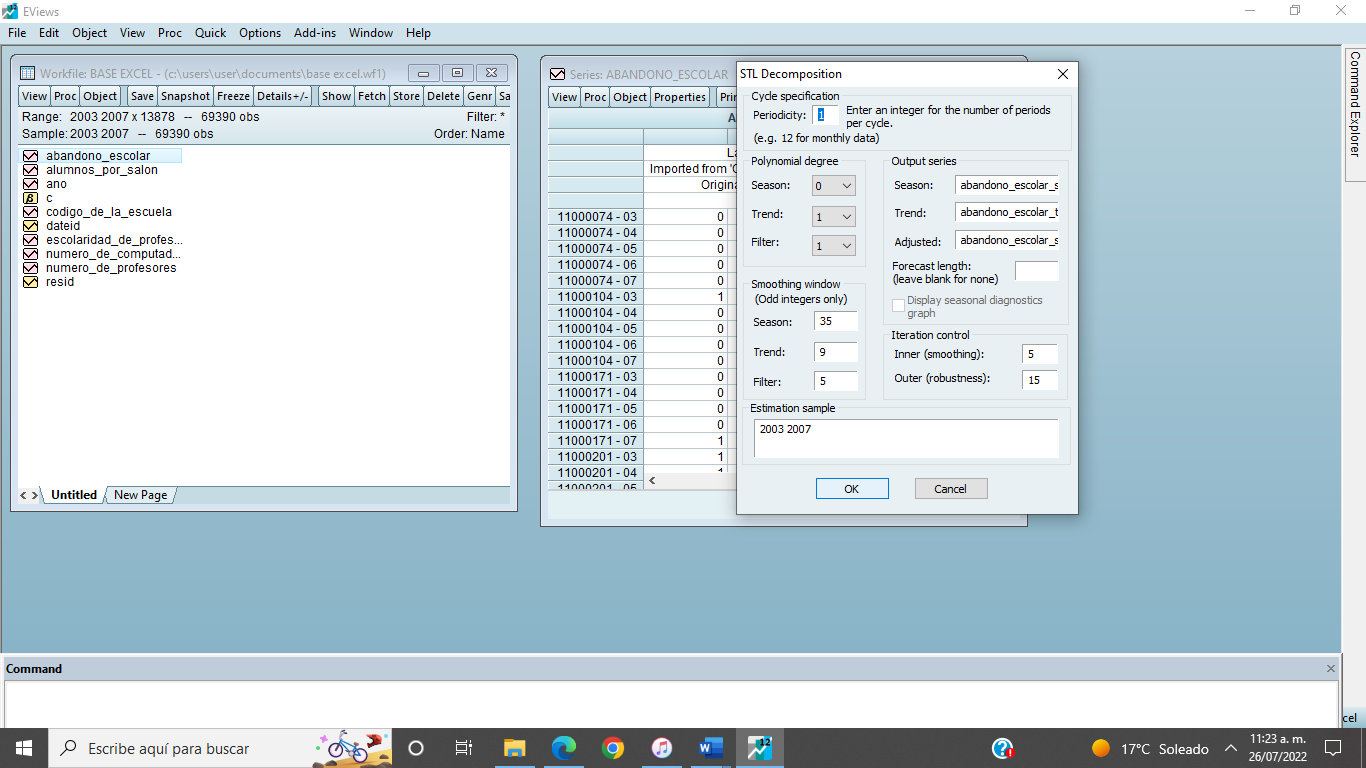
2.- Años de escolaridad de los profesores: variable discreta (8, 15; de 8 años a 15 o más).

3.- Alumnos por salón: variable discreta (1, 51; de 1 alumno a más de 50).

4.- Número de computadoras: variable discreta (0, 201; de no tener hasta tener más de 200).

**III.- Operacionalización y Análisis.**

Antes de proceder con el análisis, se utilizó el software de Eviews para desestacionalizar y ajustar las series a una periodicidad de 1:



Con esta transformación, se comienza a trabajar la serie en STATA.

Se define el panel en términos de las escuelas y los años en que se recolectaron los datos:

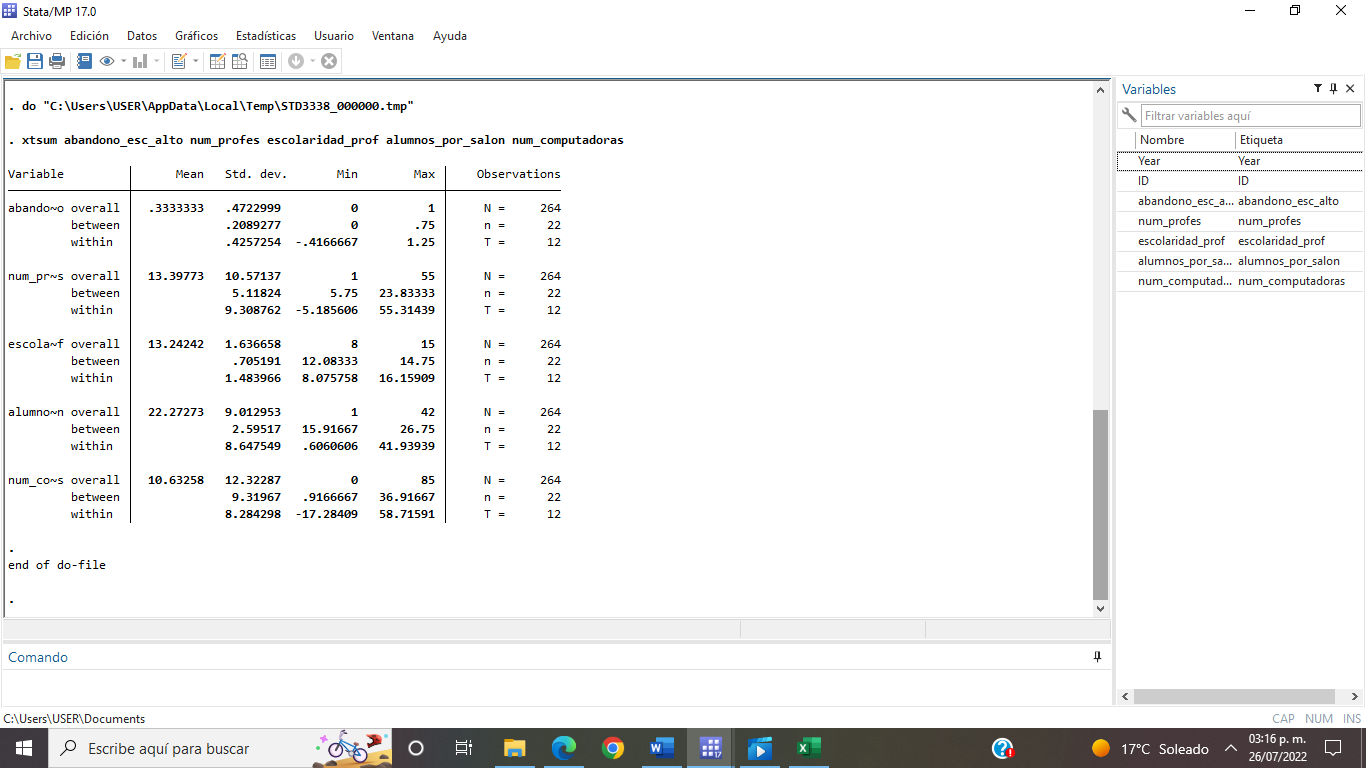
. xtset ID Year

Panel variable: ID (strongly balanced)

Time variable: Year, 2003 to 2014

Delta: 1 unit

Tras definir el panel, observamos su distribución a partir de las diferentes categorías:



Como se puede observar, las variables de **abandono escolar**, **número de profesores** y **número de computadoras** se destacan por tener medias relativamente pequeñas comparados a la gama de valores que pueden tomar. En el caso de las variables de **alumnos por salón** y **nivel de escolaridad de los profesores**, sus promedios están sustancialmente más cerca de los valores máximos. Al fijarnos en las diferencias entre las variables, observamos que algunas variables presentan cambios drásticos (**número de computadoras**, **número de profesores** y **alumnos por salón de clase**), mientras que parece haber relativamente poca diferencia entre otras (**escolaridad de los profesores**). Quizás más interesante, al observar las diferencias entre las mismas variables de una escuela a través del tiempo, es evidente que existen cambios muy bruscos en muchas de las variables (**número de profesores**, **escolaridad de los profesores**, **alumnos por salón** y **número de computadoras**). Este hallazgo es poco intuitivo, pues indica mucha volatilidad en los insumos y el capital humano que poseen las escuelas en poco más de una década, lo que podría indicar que la eficacia de una escuela podría ser difícil de mantener a través del tiempo.

1. Gráficos del Comportamiento de las escuelas:

Al observar un gráfico sobre el abandono escolar a través del tiempo en las 22 escuelas, se obtienen los siguientes diagramas:



A partir de la tabla con las gráficas, es evidente que, en la mayoría de las escuelas, hay una fluctuación importante en el abandono escolar, habiendo solamente una escuela en toda la serie que pudo mantener un nivel bajo de abandono escolar por los 12 años (‘20’). Asimismo, es posible decir que, para el inicio del periodo de tiempo de observación (2003), 15 escuelas (68%) reportaban tener un alto abandono escolar, mientras que para el último año (2014), solamente 4 (18%) del total de escuelas mantenía esta tendencia. Esta tendencia podría ofrecer algún sustento a la teoría utilizada para el ejercicio: al igual como sucede con las empresas, en un inicio, las escuelas podrían experimentar más dificultades a falta de experiencia y una posición solida en el mercado. Sin embargo, con el paso del tiempo, las escuelas adquieren más experiencia, renombre y capital (humano y físico) que les permite ofrecer un mejor servicio a sus clientes, con lo cual comienza a experimentar menos bajas y problemas de funcionamiento.

Respecto al comportamiento de las escuelas en torno a su número de profesores y de computadoras, observamos los siguientes resultados:

****

Tras observar el comportamiento de ambas variables a través del tiempo, los resultados indican sustento mixto para la teoría de la productividad de las escuelas. Por un lado, en el caso del número de computadoras que tiene una escuela, en 10 casos (45%) se puede observar que las escuelas incrementan gradualmente el número de aparatos en sus instalaciones, en 10 la cifra se mantiene estancada, mientras que en solo 2 casos parece haber una reducción en el número de ordenadores. A partir de esta tendencia, es posible pensar que las escuelas que tienen recursos buscan incrementar el número de aparatos que tienen disponibles a sus alumnos, con el fin de aumentar la calidad de la enseñanza con más recursos digitales, mientras que hay pocas que permiten que se descompongan o retiren los aparatos de cómputo.

Por otro lado, las escuelas no parecen mostrar una tendencia de incrementar sus recursos humanos en el tiempo. De toda la muestra, solo en 6 casos (27%) se puede observar una tendencia a la alza en el número de profesores con el paso del tiempo. Por el contrario, en hasta 15 escuelas (68%) hubo una disminución del número de profesores con el paso del tiempo, lo cual podría indicar un descuido en el capital humano. Quizás haya otros factores, como asuntos sindicales y restricciones al presupuesto de las escuelas, que puedan estar afectando su capacidad de incrementar su número de profesores.

Respecto al nivel de escolaridad de los profesores, se observan los siguientes resultados:



Al observar el comportamiento de la variable por escuela, se observa una fluctuación, constante, al grado que en muchos casos, los años de escolaridad que tienen los profesores se mantiene en prácticamente el mismo lugar con el paso de los años. Quizás esta situación sea producto de un tema laboral en las escuelas: los maestros podrían tener incentivos altos de abandonar sus trabajos después de un tiempo, por lo que los dirigentes de las escuelas buscan reponer a los profesores salientes con profesores con una preparación semejante.

Finalmente, en el caso del número de alumnos por salón de clase, el comportamiento de esta variable en las 22 escuelas es el siguiente:



Al observar el comportamiento de esta variable, es evidente que no se observa un patrón claro en el largo plazo. Quizás la única tendencia que se puede observar es que las escuelas tienden a iniciar con salones con muchos alumnos, logran disminuir esta cifra en el paso de los años, pero luego vuelven a posicionarse en niveles semejantes a los iniciales. Es posible que este proceso podría estar siguiendo una lógica empresarial (es decir, gradualmente aumentar el número de salones para poder ofrecer educación a más alumnos), pero quizás no se esté tomando en cuenta si esta estrategia reduce la calidad del aprendizaje de los alumnos.

1. Gráficos de dispersión bivariados:

A continuación, se llevan a cabo gráficos de dispersión entre la variable dependiente y las variables regresoras:

1.- Abandono escolar y número de profesores:



A primera vista, no se puede notar una tendencia clara entre el efecto que tiene el número de profesores sobre el abandono escolar, aunque la dispersión de la variable podría indicar que podría el número de profesores podría tener algún efecto sobre el abandono escolar.

2.- Abandono escolar y número de profesores:



A primera vista, no se puede observar la existencia de un efecto entre la escolaridad de los profesores y el abandono escolar de los alumnos.

3.- Abandono escolar y alumnos por salón de clase:

****

A primera vista, no se puede notar una tendencia clara entre el efecto que tiene el número de profesores sobre el abandono escolar. A pesar de ello, la dispersión de la variable podría indicar que hay un mayor nivel de abandono escolar cuando los salones de escuela tienen más alumnos.

4.- Abandono escolar y número de computadoras:

****

A primera vista, parece que un mayor número de computadoras en la escuela está relacionado con un menor número de abandono escolar, pero esta relación no parece muy clara ni fuerte.

1. Gráficos de valor estimado de la variable independiente:

Tras conocer el comportamiento agregado de las variables, es útil tener una imagen del comportamiento individual de las escuelas. Para eso, se elaboran gráficos del valor estimado de abandono escolar por cada escuela. Para ello, se calcula una regresión de efectos fijos, asumiendo que cada escuela es una variable dicotómica:

xi: reg abandono num\_prof escolaridad alumnos num\_compu i.ID

i.ID \_IID\_1-22 (naturally coded; \_IID\_1 omitted)

Source | SS df MS Number of obs = 264

-------------+---------------------------------- F(25, 238) = 3.91

Model | 17.0717431 25 .682869723 Prob > F = 0.0000

Residual | 41.5949236 238 .174768587 R-squared = 0.2910

-------------+---------------------------------- Adj R-squared = 0.2165

Total | 58.6666667 263 .223067174 Root MSE = .41805

---------------------------------------------------------------------------------

abandono\_esc\_alto | Coefficient Std. err. t P>|t| [95% conf. interval]

------------------+--------------------------------------------------------------

num\_profes | .0036779 .0031996 1.15 0.252 -.0026252 .0099809

escolaridad\_prof | -.0286758 .0189004 -1.52 0.131 -.0659092 .0085576

alumnos\_por\_salon | .0104568 .0032792 3.19 0.002 .0039969 .0169167

num\_computadoras | -.0095485 .003298 -2.90 0.004 -.0160454 -.0030515

\_IID\_2 | .5036992 .1743858 2.89 0.004 .1601624 .847236

\_IID\_3 | .0806087 .1735134 0.46 0.643 -.2612094 .4224268

\_IID\_4 | -.0605028 .1799623 -0.34 0.737 -.4150252 .2940197

\_IID\_5 | -.175994 .172791 -1.02 0.309 -.5163892 .1644011

\_IID\_6 | .084687 .1728227 0.49 0.625 -.2557705 .4251445

\_IID\_7 | -.0229939 .1789406 -0.13 0.898 -.3755036 .3295159

\_IID\_8 | .4440502 .1805185 2.46 0.015 .0884321 .7996683

\_IID\_9 | .0921844 .1756202 0.52 0.600 -.2537842 .438153

\_IID\_10 | .6003593 .1736378 3.46 0.001 .258296 .9424226

\_IID\_11 | .2395565 .1820163 1.32 0.189 -.1190123 .5981253

\_IID\_12 | -.1174325 .1887784 -0.62 0.534 -.4893224 .2544574

\_IID\_13 | -.0515743 .181329 -0.28 0.776 -.408789 .3056405

\_IID\_14 | .3692136 .1811529 2.04 0.043 .0123457 .7260815

\_IID\_15 | -.1493646 .1803497 -0.83 0.408 -.5046502 .205921

\_IID\_16 | -.1605928 .1763225 -0.91 0.363 -.5079448 .1867592

\_IID\_17 | .1017015 .1850162 0.55 0.583 -.262777 .46618

\_IID\_18 | .1236458 .1781814 0.69 0.488 -.2273681 .4746598

\_IID\_19 | .2703566 .1786978 1.51 0.132 -.0816747 .6223879

\_IID\_20 | -.2649467 .1892112 -1.40 0.163 -.6376892 .1077958

\_IID\_21 | .0977281 .1732205 0.56 0.573 -.2435132 .4389694

\_IID\_22 | .1348035 .1845799 0.73 0.466 -.2288156 .4984225

\_cons | .4351832 .2964131 1.47 0.143 -.1487451 1.019112

---------------------------------------------------------------------------------

A partir de ésta, es posible conocer el comportamiento de cada escuela por variable.

1.- Abandono escolar y número de profesores:



Como se puede observar, a primera vista no es posible concluir si existe una relación lineal notable entre estas variables, ni es claro cuál es su sentido.

2.- Abandono escolar y escolaridad de profesores:

****

Como se puede observar, a primera vista no es posible concluir si existe una relación lineal notable entre estas variables, ni es claro cuál es su sentido.

3.- Abandono escolar y número de alumnos por salón:

****

Como se puede observar, a primera vista no es posible concluir si existe una relación lineal notable entre estas variables, ni es claro cual es su sentido. Sin embargo, hay algunos indicios, como el sentido de la pendiente, que podrían indicar que el abandono escolar es afectado de manera positiva por el número de alumnos por salón.

4.- Abandono escolar y número de computadoras:

****

Como se puede observar, a primera vista no es posible concluir si existe una relación lineal notable entre estas variables, ni es claro cuál es su sentido. Sin embargo, hay algunos indicios, como el sentido de la pendiente, que podrían indicar que el abandono escolar es afectado de manera negativa por el número de computadoras que tiene una escuela.

En resumen, es posible observar que existe una varianza notable en los efectos de la variable para cada escuela. Sin embargo, se divisan algunas tendencias que podrían indicar que, más allá de casos individuales, las variables podrían mostrar tendencias generalizadas.

Con el fin de conocer la distribución de cada una de las escuelas, se elabora un gráfico de la media del abandono escolar de cada una de las 22 escuelas.



Como se puede observar, existe una varianza notable entre la media de abandono escolar de cada escuela. Mientras que solamente una escuela nunca se halló en una situación de alto abandono escolar (‘20’), dos escuelas (‘8’ y ‘10’) se mantuvieron en una situación de abandono escolar durante gran parte del periodo de tiempo (promedio de más de 0.75).

También se elabora un gráfico de la media del abandono escolar durante los 12 años de estudio:



Como se puede observar, hay una tendencia de alto abandono escolar entre las escuelas durante los primeros 4 años, la cual cae notablemente y, para los últimos 4 años, permanece estable.

A partir de esta varianza en las observaciones de las escuelas y durante el periodo de tiempo, parece que será posible elaborar un modelo con estimación de regresión del fenómeno.

Finalmente, se procede a hacer la estadística descriptiva de los datos de panel.

. xtdescribe

ID: 1, 2, ..., 22 n = 22

Year: 2003, 2004, ..., 2014 T = 12

Delta(Year) = 1 unit

Span(Year) = 12 periods

(ID\*Year uniquely identifies each observation)

Distribution of T\_i: min 5% 25% 50% 75% 95% max

12 12 12 12 12 12 12

Freq. Percent Cum. | Pattern

---------------------------+--------------

22 100.00 100.00 | 111111111111

---------------------------+--------------

22 100.00 | XXXXXXXXXXXX

Como se puede observar, el panel consiste de 22 observaciones (escuelas) en 12 años (2003 - 2014), por lo que se tiene un total de 264 observaciones. Al mostrar la distribución temporal por los diferentes cuantiles (mínimo, 5%, cuartil 1, mediana, cuartil 3, 95% y máximo), se observa que cada uno de éstos contiene exactamente 12 observaciones, lo que implica que el panel está perfectamente balanceado.

A partir de estos datos, de forma preliminar, parecería que es factible utilizar esta base de datos mediante la metodología panel, pues contiene suficientes observaciones de las variables para asemejarse a su distribución normal, al igual que contiene una cantidad lo suficientemente reducida de elementos a observar (escuelas) que hacen factible crear un panel de efectos fijos mediante variables binarias. Sin embargo, se realizarán numerosas pruebas para tener la certeza cuantitativa que es factible llevar a cabo un modelo panel para esta base de datos.

1. Modelos Panel Logísticos (estáticos).

Para conocer si las variables relacionadas con los atributos de una escuela tienen un efecto destacado sobre la variable, es necesario elaborar una serie de modelos de regresión que sirvan para conocer la naturaleza y tipo de efecto de los modelos de panel. Con este fin, se llevan a cabo 4 modelos de regresión pertinentes a la base de datos: un modelo Agrupado, de Efectos Fijos, de Efectos Fijos con variables binarias y de Efectos Aleatorios. Todos los modelos empleados utilizan la metodología de la regresión logística, ya que se espera que un modelo probabilístico ofrezca resultados más claros del impacto sobre la variable dependiente. Asimismo, a diferencia del Modelo de Probabilidad Lineal, los resultados no violan los supuestos de la distribución de probabilidad, lo cual favorece su validez y aplicación práctica.

1.- Modelo Agrupado.

Al emplear el modelo agrupado, se lleva a cabo un modelo de estimación de regresión logística con la variable dependiente y las 4 independientes. El resultado es el siguiente:

logit abandono\_esc\_alto num\_profes escolaridad\_prof alumnos\_por\_salon num\_compu

Iteration 0: log likelihood = -168.03974

Iteration 1: log likelihood = -154.24267

Iteration 2: log likelihood = -153.9998

Iteration 3: log likelihood = -153.99969

Iteration 4: log likelihood = -153.99969

Logistic regression Number of obs = 264

LR chi2(4) = 28.08

Prob > chi2 = 0.0000

Log likelihood = -153.99969 Pseudo R2 = 0.0836

-------------------------------------------------------------------------------

abandono\_esc\_alto | Coefficient Std. err. z P>|z| [95% conf. interval]

------------------+------------------------------------------------------------

num\_profes | .0160519 .0152264 1.05 0.292 -.0137913 .0458951

escolaridad\_prof | -.1762332 .0967843 -1.82 0.069 -.3659271 .0134606

alumnos\_por\_salon | .0724241 .0184014 3.94 0.000 .036358 .1084901

num\_computadoras | -.01272 .0129129 -0.99 0.325 -.0380288 .0125888

\_cons | -.1448893 1.224908 -0.12 0.906 -2.545665 2.255886

-------------------------------------------------------------------------------

A partir de los resultados, a excepción del número de profesores por escuela, las variables se comportan de la forma esperada. De esta forma, el **número de profesores** parece tener un efecto positivo sobre el abandono escolar, aunque éste no es estadísticamente significativo ni siquiera a un nivel de significancia de 10%. En el caso de la **escolaridad de los profesores**, esta variable tiene un efecto negativo esperado y tiene un nivel de significancia mayor al 7%, lo cual apunta a que esta variable tiene un efecto negativo relevante sobre el abandono escolar. La variable que parece tener el mayor impacto sobre el abandono escolar es el n**úmero de alumnos por salón**, pues muestra el efecto positivo esperado y es estadísticamente significativo a un nivel de significancia mayor a 0.001%, lo cual hace pensar que esta variable tiene un efecto notable sobre el abandono escolar. Finalmente, la variable de **número de computadoras** tiene el efecto negativo esperado sobre el abandono escolar, pero este no parece ser estadísticamente significativo (ni siquiera alcanza un intervalo de confianza mayor a 90%). La constante tampoco pasa la prueba de significancia estadística.

Respecto a la pseudo R2, esta muestra un valor ligeramente mayor a 0.08, indicando que el porcentaje total de valores correctamente clasificados contra los incorrectos es solamente poco mayor a 8%.

A continuación, se observan la razón de productos cruzados (Odds Ratio) de cada variable:

Logistic regression Number of obs = 264

LR chi2(4) = 28.08

Prob > chi2 = 0.0000

Log likelihood = -153.99969 Pseudo R2 = 0.0836

-------------------------------------------------------------------------------

abandono\_esc\_alto | Odds ratio Std. err. z P>|z| [95% conf. interval]

------------------+------------------------------------------------------------

num\_profes | 1.016181 .0154728 1.05 0.292 .9863033 1.046965

escolaridad\_prof | .8384224 .0811462 -1.82 0.069 .6935534 1.013552

alumnos\_por\_salon | 1.075111 .0197835 3.94 0.000 1.037027 1.114594

num\_computadoras | .9873606 .0127497 -0.99 0.325 .9626852 1.012668

\_cons | .8651181 1.05969 -0.12 0.906 .0784209 9.543749

-------------------------------------------------------------------------------

Note: \_cons estimates baseline odds.

A partir de los resultados, es posible concluir que la variable que tiene el mayor impacto sobre el abandono escolar es el número de alumnos por salón (1.07), seguido por el número de profesores (1.01). Asimismo, la escolaridad de los profesores parece ser la que tiene el menor impacto sobre el abandono escolar (0.83).

En el caso de los productos cruzados (Odds Ratio) del modelo, se obtienen los siguientes resultados:

. display abandono\_esc\_alto\_p / q

.34018716

El resultado indica que las cuatro variables principales del modelo: el número de profesores, la escolaridad de éstos, el número de alumnos por salón y computadoras inciden en un 0.34 en que una escuela experimente un alto abandono escolar.

Finalmente, se obtienen los efectos marginales del modelo:

. margins, dydx(\*)

Average marginal effects Number of obs = 264

Model VCE: OIM

Expression: Pr(abandono\_esc\_alto), predict()

dy/dx wrt: num\_profes escolaridad\_prof alumnos\_por\_salon num\_computadoras

-------------------------------------------------------------------------------

| Delta-method

| dy/dx std. err. z P>|z| [95% conf. interval]

------------------+------------------------------------------------------------

num\_profes | .0032021 .0030128 1.06 0.288 -.002703 .0091071

escolaridad\_prof | -.0351554 .018851 -1.86 0.062 -.0721027 .0017919

alumnos\_por\_salon | .0144473 .0032906 4.39 0.000 .0079979 .0208967

num\_computadoras | -.0025374 .0025608 -0.99 0.322 -.0075566 .0024818

Como se puede observar, solamente las variables de alumnos por salón y escolaridad de los profesores parecen tener efectos marginales estadísticamente significativos. En este sentido, un aumento en la escolaridad de los profesores resulta en una reducción de 3% en la probabilidad de que la escuela se hallé en una situación de alto abandono escolar, a un nivel de significancia de cerca del 6%. En el caso del número de alumnos por salón, un aumento de 1 alumno incrementa en 1.4% la probabilidad de que la escuela se encuentre en una situación de alto abandono escolar, a un nivel de significancia estadística mayor al 0.001%. El efecto del número de profesores y computadoras son muy pequeños y no poseen un nivel de significancia estadística mayor siquiera al 10%.

En general, es posible observar que el modelo contribuye a explicar porqué una escuela se encuentra en una situación de alto abandono escolar, pero algunas de las variables importantes no son estadísticamente significativas y la eficiencia del modelo podría ser mayor.

2.- Modelo de Efectos Aleatorios.

Se emplea un modelo panel de estimación de regresión logística con la variable dependiente y las 4 independientes por efectos aleatorios. El resultado es el siguiente:

. xtlogit abandono\_esc\_alto num\_profes escolaridad\_prof alumnos\_por\_salon num\_compu, re

Fitting comparison model:

Iteration 0: log likelihood = -168.03974

Iteration 1: log likelihood = -154.24267

Iteration 2: log likelihood = -153.9998

Iteration 3: log likelihood = -153.99969

Iteration 4: log likelihood = -153.99969

Fitting full model:

tau = 0.0 log likelihood = -153.99969

tau = 0.1 log likelihood = -151.30526

tau = 0.2 log likelihood = -149.62776

tau = 0.3 log likelihood = -148.68688

tau = 0.4 log likelihood = -148.34062

tau = 0.5 log likelihood = -148.54306

Iteration 0: log likelihood = -148.34036

Iteration 1: log likelihood = -147.58651

Iteration 2: log likelihood = -147.58375

Iteration 3: log likelihood = -147.58375

Random-effects logistic regression Number of obs = 264

Group variable: ID Number of groups = 22

Random effects u\_i ~ Gaussian Obs per group:

min = 12

avg = 12.0

max = 12

Integration method: mvaghermite Integration pts. = 12

Wald chi2(4) = 24.37

Log likelihood = -147.58375 Prob > chi2 = 0.0001

-------------------------------------------------------------------------------

abandono\_esc\_alto | Coefficient Std. err. z P>|z| [95% conf. interval]

------------------+------------------------------------------------------------

num\_profes | .0208071 .0174907 1.19 0.234 -.0134741 .0550883

escolaridad\_prof | -.1707759 .1091367 -1.56 0.118 -.3846799 .0431281

alumnos\_por\_salon | .0693982 .0198859 3.49 0.000 .0304226 .1083738

num\_computadoras | -.0324234 .0186547 -1.74 0.082 -.068986 .0041393

\_cons | -.1188355 1.397487 -0.09 0.932 -2.857859 2.620188

------------------+------------------------------------------------------------

/lnsig2u | -.120745 .5609138 -1.220116 .9786259

------------------+------------------------------------------------------------

sigma\_u | .9414138 .264026 .5433194 1.631195

rho | .2122205 .0937752 .0823405 .4471428

-------------------------------------------------------------------------------

LR test of rho=0: chibar2(01) = 12.83 Prob >= chibar2 = 0.000

A partir de los resultados, la probabilidad chi-cuadrada indica que la regresión es valida. Asimismo, se observan algunos patrones semejantes al modelo agrupado. En primer lugar, las variables mantuvieron efectos de la misma dirección que en el modelo anterior, por lo que solamente la variable de **número de profesores** se comportó de forma contraria a lo esperado. Respecto a la significancia estadística de las variables, la única variable que muestra una significancia estadística mayor a un nivel de 0.001 es el **número de alumnos por salón**, lo que indica que es probable que, a mayor número de alumnos por salón tenga una escuela, más probable es que sufra de una situación de abandono escolar alto. Otra variable que tiene una significancia estadística notable es el **número de computadoras** que tiene la escuela, la cual tiene el efecto negativo esperado y alcanza una significancia mayor a 9%. Esta cifra indica que el número de computadoras que tiene una escuela podría reducir la probabilidad de que presente un abandono escolar alto. La constante es negativa y no es estadísticamente significativa.

Asimismo, la cifra de .2122 de rho indica el porcentaje de la varianza que se debe al término individual, lo cual muestra que la parte individual ocupa una parte relativamente pequeña.

A continuación, se observan la razón de productos cruzados (Odds Ratio) de cada variable:

. xtlogit abandono\_esc\_alto num\_profes escolaridad\_prof alumnos\_por\_salon num\_compu, or

Fitting comparison model:

Iteration 0: log likelihood = -168.03974

Iteration 1: log likelihood = -154.24267

Iteration 2: log likelihood = -153.9998

Iteration 3: log likelihood = -153.99969

Iteration 4: log likelihood = -153.99969

Fitting full model:

tau = 0.0 log likelihood = -153.99969

tau = 0.1 log likelihood = -151.30526

tau = 0.2 log likelihood = -149.62776

tau = 0.3 log likelihood = -148.68688

tau = 0.4 log likelihood = -148.34062

tau = 0.5 log likelihood = -148.54306

Iteration 0: log likelihood = -148.34036

Iteration 1: log likelihood = -147.58651

Iteration 2: log likelihood = -147.58375

Iteration 3: log likelihood = -147.58375

Random-effects logistic regression Number of obs = 264

Group variable: ID Number of groups = 22

Random effects u\_i ~ Gaussian Obs per group:

min = 12

avg = 12.0

max = 12

Integration method: mvaghermite Integration pts. = 12

Wald chi2(4) = 24.37

Log likelihood = -147.58375 Prob > chi2 = 0.0001

-------------------------------------------------------------------------------

abandono\_esc\_alto | Odds ratio Std. err. z P>|z| [95% conf. interval]

------------------+------------------------------------------------------------

num\_profes | 1.021025 .0178585 1.19 0.234 .9866163 1.056634

escolaridad\_prof | .8430105 .0920034 -1.56 0.118 .6806685 1.044072

alumnos\_por\_salon | 1.071863 .0213149 3.49 0.000 1.03089 1.114464

num\_computadoras | .9680966 .0180596 -1.74 0.082 .9333398 1.004148

\_cons | .8879538 1.240904 -0.09 0.932 .0573915 13.73831

------------------+------------------------------------------------------------

/lnsig2u | -.120745 .5609138 -1.220116 .9786259

------------------+------------------------------------------------------------

sigma\_u | .9414138 .264026 .5433194 1.631195

rho | .2122205 .0937752 .0823405 .4471428

-------------------------------------------------------------------------------

Note: Estimates are transformed only in the first equation to odds ratios.

LR test of rho=0: chibar2(01) = 12.83 Prob >= chibar2 = 0.000

Los resultados son prácticamente iguales a los del modelo agrupado: la variable que tiene el mayor impacto sobre el abandono escolar es el número de alumnos por salón (1.07), seguido por el número de profesores (1.02). Asimismo, la escolaridad de los profesores parece ser la que tiene el menor impacto sobre el abandono escolar (0.84).

En el caso del producto cruzado (Odds Ratio) del modelo, se asume que es muy semejante al del modelo agrupado: el número de profesores, la escolaridad de éstos, el número de alumnos por salón y computadoras inciden en un 0.34 en que una escuela experimente un alto abandono escolar.

Finalmente, se obtienen los efectos marginales del modelo:

. margins, dydx(\*)

Average marginal effects Number of obs = 264

Model VCE: OIM

Expression: Pr(abandono\_esc\_alto=1), predict(pr)

dy/dx wrt: num\_profes escolaridad\_prof alumnos\_por\_salon num\_computadoras

-------------------------------------------------------------------------------

| Delta-method

| dy/dx std. err. z P>|z| [95% conf. interval]

------------------+------------------------------------------------------------

num\_profes | .0035982 .0029978 1.20 0.230 -.0022773 .0094737

escolaridad\_prof | -.0295322 .0186656 -1.58 0.114 -.0661161 .0070517

alumnos\_por\_salon | .012001 .0033363 3.60 0.000 .005462 .01854

num\_computadoras | -.005607 .0030854 -1.82 0.069 -.0116542 .0004403

-------------------------------------------------------------------------------

Como se puede observar, solamente las variables de **alumnos por salón** y el **número de computadoras** parecen tener efectos marginales estadísticamente significativos. En este sentido, en el número de alumnos por salón, un aumento de 1 alumno incrementa en 1.2% la probabilidad de que la escuela se encuentre en una situación de alto abandono escolar, a un nivel de significancia estadística mayor al 0.001%. En el caso del número de computadoras, cada computadora adicional que tenga la escuela reduce en 0.5% la probabilidad de que la escuela experimente un alto abandono escolar, a un nivel de significancia superior a 7%. El efecto del número de profesores y escolaridad de los profesores son muy pequeños y no poseen un nivel de significancia estadística mayor siquiera al 10%.

En general, es posible observar que este modelo también contribuye a explicar por qué una escuela se encuentra en una situación de alto abandono escolar. Sin embargo, algunas de las variables importantes no son estadísticamente significativas.

3.- Modelo de Efectos Fijos.

Se emplea un modelo panel de estimación de regresión logística con la variable dependiente y las 4 independientes por efectos fijos. El resultado es el siguiente:

. xtlogit abandono\_esc\_alto num\_profes escolaridad\_prof alumnos\_por\_salon num\_compu, fe

note: multiple positive outcomes within groups encountered.

note: 1 group (12 obs) omitted because of all positive or

all negative outcomes.

Iteration 0: log likelihood = -99.275784

Iteration 1: log likelihood = -97.223026

Iteration 2: log likelihood = -97.205867

Iteration 3: log likelihood = -97.205862

Conditional fixed-effects logistic regression Number of obs = 252

Group variable: ID Number of groups = 21

Obs per group:

min = 12

avg = 12.0

max = 12

LR chi2(4) = 31.23

Log likelihood = -97.205862 Prob > chi2 = 0.0000

-------------------------------------------------------------------------------

abandono\_esc\_alto | Coefficient Std. err. z P>|z| [95% conf. interval]

------------------+------------------------------------------------------------

num\_profes | .01644 .018691 0.88 0.379 -.0201937 .0530738

escolaridad\_prof | -.1475758 .1142867 -1.29 0.197 -.3715736 .076422

alumnos\_por\_salon | .0560011 .0198687 2.82 0.005 .0170592 .0949431

num\_computado | -.0613118 .0241715 -2.54 0.011 -.1086871 -.01393

-------------------------------------------------------------------------------

Como se puede observar, el programa eliminó una de las escuelas del modelo por tener efectos estáticos durante los 12 años. A pesar de ello, la probabilidad de la prueba chi cuadrada muestra que la regresión es válida, por lo que se continúa con el análisis. Al igual que en el caso del modelo por efectos fijos, el **número de alumnos por salón** (efecto positivo) y el **número de computadoras** (efecto negativo) son las variables más relevantes al abandono escolar, ya que tienen un nivel de significancia mayor a 5%. Los efectos del número y escolaridad de los profesores no alcanzan una relevancia estadística mínima (10%, 5% o del 1%).

Al analizar el producto cruzado de la regresión (Odds Ratios), se obtiene el siguiente resultado:

Fitting comparison model:

Iteration 0: log likelihood = -163.0324

Iteration 1: log likelihood = -149.97768

Iteration 2: log likelihood = -149.78412

Iteration 3: log likelihood = -149.78404

Iteration 4: log likelihood = -149.78404

Fitting full model:

tau = 0.0 log likelihood = -149.78404

tau = 0.1 log likelihood = -147.48999

tau = 0.2 log likelihood = -146.14878

tau = 0.3 log likelihood = -145.50093

tau = 0.4 log likelihood = -145.41944

tau = 0.5 log likelihood = -145.86931

Iteration 0: log likelihood = -145.41934

Iteration 1: log likelihood = -144.81884

Iteration 2: log likelihood = -144.81333

Iteration 3: log likelihood = -144.81333

Random-effects logistic regression Number of obs = 252

Group variable: ID Number of groups = 21

Random effects u\_i ~ Gaussian Obs per group:

min = 12

avg = 12.0

max = 12

Integration method: mvaghermite Integration pts. = 12

Wald chi2(4) = 23.58

Log likelihood = -144.81333 Prob > chi2 = 0.0001

-------------------------------------------------------------------------------

abandono\_esc\_alto | Odds ratio Std. err. z P>|z| [95% conf. interval]

------------------+------------------------------------------------------------

num\_profes | 1.019365 .0176102 1.11 0.267 .9854276 1.054472

escolaridad\_prof | .8348769 .0907551 -1.66 0.097 .6746721 1.033123

alumnos\_por\_salon | 1.068961 .0211205 3.38 0.001 1.028357 1.111168

num\_computadoras | .9668035 .0175956 -1.85 0.064 .9329247 1.001913

\_cons | 1.239643 1.725587 0.15 0.877 .0809882 18.97455

------------------+------------------------------------------------------------

/lnsig2u | -.3362208 .5927546 -1.497998 .825557

------------------+------------------------------------------------------------

sigma\_u | .8452605 .250516 .4728395 1.51101

rho | .178423 .0868909 .0636347 .4096795

-------------------------------------------------------------------------------

Note: Estimates are transformed only in the first equation to odds ratios.

Note: \_cons estimates baseline odds (conditional on zero random effects).

LR test of rho=0: chibar2(01) = 9.94 Prob >= chibar2 = 0.001

De acuerdo a los resultados, se mantiene el patrón de los modelos anteriores: la variable de alumnos por salón es la que tiene mayor importancia sobre el abandono de alumnos, seguido del número de profesores, el número de computadoras y, finalmente, la escolaridad de los profesores. A pesar de ello, hay 2 hechos interesantes del modelo. En primer lugar, la variable de escolaridad de profesores y número de computadoras ambas alcanzan una significancia estadística mayor al 10%, lo cual no se había alcanzado en modelos anteriores. En segundo lugar, el valor de rho es menor al modelo anterior (de .22 a .17) lo cual indica que el error debido a la parte individual pudo haber disminuido al aplicar efectos fijos.

En el caso del producto cruzado (Odds Ratio) del modelo, también se asume que es muy semejante al del modelo agrupado: el número de profesores, la escolaridad de éstos, el número de alumnos por salón y computadoras inciden en un 0.34 en que una escuela experimente un alto abandono escolar.

Finalmente, se obtienen los efectos marginales:

. margins, dydx(\*)

Average marginal effects Number of obs = 252

Model VCE: OIM

Expression: Pr(abandono\_esc\_alto|fixed effect is 0), predict(pu0)

dy/dx wrt: num\_profes escolaridad\_prof alumnos\_por\_salon num\_computadoras

-------------------------------------------------------------------------------

| Delta-method

| dy/dx std. err. z P>|z| [95% conf. interval]

------------------+------------------------------------------------------------

num\_profes | .0029005 .0031509 0.92 0.357 -.003275 .0090761

escolaridad\_prof | -.0260369 .009549 -2.73 0.006 -.044752 -.007321

alumnos\_por\_salon | .0098803 .0064999 1.52 0.128 -.0028593 .0226199

num\_computadoras | -.0108173 .0065556 -1.65 0.099 -.0236661 .0020315

-------------------------------------------------------------------------------

En esta ocasión, es posible observar que, contrario a lo que se divisa en la regresión general, cambia la significancia estadística del efecto marginal de las variables: la escolaridad de los profesores es la única variable que alcanza una significancia estadística mayor a 5%, seguida del número de computadoras con una significancia mayor a 10%, mientras que la variable de alumnos por salón de clase no alcanza una significancia estadística de 10%, 5% o 1%. En esta situación, por cada año de escolaridad del profesor, se reduce en un 2% la probabilidad de que la escuela tenga un problema de alto abandono escolar. Respecto al número de computadoras, por cada computadora adicional que tenga la escuela, se reduce en un 1% la probabilidad de que la escuela experimente una situación de alto abandono escolar. Estos efectos inesperados sobre el comportamiento de las variables podrían deberse a que un modelo de efectos fijos afecta la estimación de predictores que no cambian a través del tiempo, alterando los resultados del modelo.

A final de cuentas, el modelo también es efectivo en explicar el efecto que tienen las variables internas de una escuela sobre el abandono escolar que enfrentan sus alumnos. A pesar de tener algunas ventajas sobre el modelo de efectos variables, como un componente de error individual más reducido, también posee desventajas notables (la inconsistencia con los efectos marginales del modelo y el modelo de regresión principal).

4.- Modelo de Efectos Fijos con Variables Binarias.

Finalmente, se estima un modelo logístico de efectos fijos para analizar el abandono escolar a partir de las cuatro variables regresoras, al igual que controlando los efectos de una de las variables de la regresión. Los resultados son los siguientes:

xi: logit abandono num\_prof escolaridad alumnos num\_compu i.ID

i.ID \_IID\_1-22 (naturally coded; \_IID\_1 omitted)

note: \_IID\_20 != 0 predicts failure perfectly;

\_IID\_20 omitted and 12 obs not used.

Iteration 0: log likelihood = -163.0324

Iteration 1: log likelihood = -125.32015

Iteration 2: log likelihood = -123.66422

Iteration 3: log likelihood = -123.6468

Iteration 4: log likelihood = -123.64679

Iteration 5: log likelihood = -123.64679

Logistic regression Number of obs = 252

LR chi2(24) = 78.77

Prob > chi2 = 0.0000

Log likelihood = -123.64679 Pseudo R2 = 0.2416

-------------------------------------------------------------------------------

abandono\_esc\_alto | Coefficient Std. err. z P>|z| [95% conf. interval]

------------------+------------------------------------------------------------

num\_profes | .0183987 .019626 0.94 0.349 -.0200676 .056865

escolaridad\_prof | -.1631157 .1200708 -1.36 0.174 -.39845 .0722187

alumnos\_por\_salon | .061652 .020877 2.95 0.003 .0207337 .1025703

num\_computadoras | -.0682525 .0256443 -2.66 0.008 -.1185145 -.01799

\_IID\_2 | 2.926252 1.179454 2.48 0.013 .6145649 5.237938

\_IID\_3 | .2056893 1.08229 0.19 0.849 -1.91556 2.326939

\_IID\_4 | -.7116494 1.142009 -0.62 0.533 -2.949945 1.526647

\_IID\_5 | -1.818231 1.380353 -1.32 0.188 -4.523673 .8872107

\_IID\_6 | .2926159 1.03642 0.28 0.778 -1.73873 2.323962

\_IID\_7 | -.56903 1.203673 -0.47 0.636 -2.928186 1.790126

\_IID\_8 | 1.87532 1.105905 1.70 0.090 -.2922137 4.042854

\_IID\_9 | .3141138 1.038094 0.30 0.762 -1.720513 2.34874

\_IID\_10 | 3.012544 1.06448 2.83 0.005 .926201 5.098886

\_IID\_11 | .9104328 1.060233 0.86 0.391 -1.167586 2.988452

\_IID\_12 | -1.045737 1.232669 -0.85 0.396 -3.461723 1.370249

\_IID\_13 | -.5495606 1.118197 -0.49 0.623 -2.741186 1.642065

\_IID\_14 | 1.892712 1.149679 1.65 0.100 -.3606169 4.146041

\_IID\_15 | -1.39878 1.369044 -1.02 0.307 -4.082057 1.284498

\_IID\_16 | -1.624591 1.359612 -1.19 0.232 -4.289382 1.0402

\_IID\_17 | .1592493 1.103998 0.14 0.885 -2.004546 2.323045

\_IID\_18 | .3610965 1.063491 0.34 0.734 -1.723308 2.445501

\_IID\_19 | 1.016198 1.072978 0.95 0.344 -1.0868 3.119196

\_IID\_20 | 0 (omitted)

\_IID\_21 | .4661261 1.050352 0.44 0.657 -1.592526 2.524778

\_IID\_22 | .4016253 1.084895 0.37 0.711 -1.72473 2.527981

\_cons | .1295007 1.847385 0.07 0.944 -3.491308 3.750309

-------------------------------------------------------------------------------

De acuerdo con estos resultados, el valor de la chi-cuadrada indica que los cálculos de la regresión son confiables. Nuevamente, el efecto de las variables de **alumnos por salón** (positivo) y **número de computadores por escuela** (negativo) son las que alcanzan un nivel de significancia mayor a 99%. En el caso de la escolaridad y número de profesores, solo la variable de número de profesores se comporta contario a la predicción del texto, pero ambas variables no alcanzan una significancia estadística mayor al 10%, 5% o 1%. Controlando por los efectos específicos de cada una de las escuelas, solamente la escuela número 2, 8, 10 y 14, alcanzaron un nivel se significancia del 10% o mayor; el resto no tuvo un impacto relevante sobre el abandono escolar. La constante es positiva, con un valor de 0.129 y un nivel de significancia estadística mayor al 10%.

Respecto a la pseudo R2, esta muestra un valor ligeramente mayor a 0.24, indicando que el porcentaje total de valores correctamente clasificados contra los incorrectos es solamente poco mayor a 24%.

Al revisar los productos cruzados (Odds Ratios) de las variables, resulta en lo siguiente:

. xi: logit abandono num\_prof escolaridad alumnos num\_compu i.ID, or

i.ID \_IID\_1-22 (naturally coded; \_IID\_1 omitted)

note: \_IID\_20 != 0 predicts failure perfectly;

\_IID\_20 omitted and 12 obs not used.

Iteration 0: log likelihood = -163.0324

Iteration 1: log likelihood = -125.32015

Iteration 2: log likelihood = -123.66422

Iteration 3: log likelihood = -123.6468

Iteration 4: log likelihood = -123.64679

Iteration 5: log likelihood = -123.64679

Logistic regression Number of obs = 252

LR chi2(24) = 78.77

Prob > chi2 = 0.0000

Log likelihood = -123.64679 Pseudo R2 = 0.2416

-------------------------------------------------------------------------------

abandono\_esc\_alto | Odds ratio Std. err. z P>|z| [95% conf. interval]

------------------+------------------------------------------------------------

num\_profes | 1.018569 .0199904 0.94 0.349 .9801325 1.058513

escolaridad\_prof | .8494929 .1019993 -1.36 0.174 .6713598 1.07489

alumnos\_por\_salon | 1.063592 .0222047 2.95 0.003 1.02095 1.108015

num\_computadoras | .9340246 .0239524 -2.66 0.008 .8882389 .9821704

\_IID\_2 | 18.65756 22.00573 2.48 0.013 1.848852 188.2815

\_IID\_3 | 1.228372 1.329454 0.19 0.849 .1472593 10.24653

\_IID\_4 | .4908339 .5605366 -0.62 0.533 .0523426 4.602716

\_IID\_5 | .1623126 .2240486 -1.32 0.188 .0108491 2.428347

\_IID\_6 | 1.339928 1.388728 0.28 0.778 .1757434 10.21607

\_IID\_7 | .5660743 .6813683 -0.47 0.636 .053494 5.990205

\_IID\_8 | 6.522906 7.213714 1.70 0.090 .7466089 56.98874

\_IID\_9 | 1.369045 1.421198 0.30 0.762 .1789744 10.47237

\_IID\_10 | 20.33907 21.65053 2.83 0.005 2.524899 163.8393

\_IID\_11 | 2.485398 2.635101 0.86 0.391 .311117 19.85492

\_IID\_12 | .3514327 .4332001 -0.85 0.396 .0313756 3.936332

\_IID\_13 | .5772034 .6454269 -0.49 0.623 .0644938 5.165824

\_IID\_14 | 6.637344 7.630812 1.65 0.100 .6972461 63.18333

\_IID\_15 | .246898 .3380143 -1.02 0.307 .0168727 3.612853

\_IID\_16 | .1969922 .2678331 -1.19 0.232 .0137134 2.829784

\_IID\_17 | 1.17263 1.294581 0.14 0.885 .1347214 10.2067

\_IID\_18 | 1.434902 1.526006 0.34 0.734 .1784747 11.53633

\_IID\_19 | 2.762672 2.964286 0.95 0.344 .3372942 22.62819

\_IID\_20 | 1 (omitted)

\_IID\_21 | 1.593808 1.674059 0.44 0.657 .2034112 12.48812

\_IID\_22 | 1.494251 1.621106 0.37 0.711 .1782212 12.52818

\_cons | 1.13826 2.102805 0.07 0.944 .030461 42.53424

-------------------------------------------------------------------------------

Note: \_cons estimates baseline odds.

A partir de los resultados, se puede observar que, entre las 4 variables regresoras principales, el número de alumnos por salón es la que tiene el efecto más sustantivo sobre el abandono escolar, mientras que la escolaridad de los profesores tiene el efecto más bajo. Respecto a las variables dicotómicas de las escuelas, la segunda escuela tiene un impacto sobresaliente con respecto al resto (18.65), mientras que las escuelas número 8 y 14 también tienen un impacto inusualmente alto (mayor a 6.0).

En el caso del producto cruzado (Odds Ratio) del modelo, también se asume que es muy semejante al del modelo agrupado: el número de profesores, la escolaridad de éstos, el número de alumnos por salón y computadoras inciden en un 0.34 en que una escuela experimente un alto abandono escolar.

Finalmente, respecto a los efectos marginales de este modelo, se obtiene el siguiente resultado:

. margins, dydx(\*)

Average marginal effects Number of obs = 252

Model VCE: OIM

Expression: Pr(abandono\_esc\_alto), predict()

dy/dx wrt: num\_profes escolaridad\_prof alumnos\_por\_salon num\_computadoras \_IID\_2 \_IID\_3 \_IID\_4 \_IID\_5 \_IID\_6 \_IID\_7 \_IID\_8 \_IID\_9 \_IID\_10 \_IID\_11

\_IID\_12 \_IID\_13 \_IID\_14 \_IID\_15 \_IID\_16 \_IID\_17 \_IID\_18 \_IID\_19 \_IID\_20 \_IID\_21 \_IID\_22

-------------------------------------------------------------------------------

| Delta-method

| dy/dx std. err. z P>|z| [95% conf. interval]

------------------+------------------------------------------------------------

num\_profes | .0029744 .0031508 0.94 0.345 -.0032011 .0091498

escolaridad\_prof | -.0263695 .0191437 -1.38 0.168 -.0638905 .0111515

alumnos\_por\_salon | .0099668 .0031614 3.15 0.002 .0037706 .0161629

num\_computadoras | -.0110338 .0039637 -2.78 0.005 -.01880 -.0032651

\_IID\_2 | .473062 .1828893 2.59 0.010 .1146056 .8315184

\_IID\_3 | .033252 .1748855 0.19 0.849 -.3095173 .3760213

\_IID\_4 | -.1150463 .1842125 -0.62 0.532 -.4760961 .2460036

\_IID\_5 | -.2939379 .2211194 -1.33 0.184 -.727324 .1394483

\_IID\_6 | .0473047 .1674356 0.28 0.778 -.2808631 .3754725

\_IID\_7 | -.0919902 .1943818 -0.47 0.636 -.4729715 .2889911

\_IID\_8 | .3031669 .1746563 1.74 0.083 -.0391531 .645487

\_IID\_9 | .0507801 .1677035 0.30 0.762 -.2779127 .3794728

\_IID\_10 | .4870121 .1616532 3.01 0.003 .1701776 .8038466

\_IID\_11 | .1471819 .1703288 0.86 0.388 -.1866565 .4810202

\_IID\_12 | -.1690553 .1984696 -0.85 0.394 -.5580486 .219938

\_IID\_13 | -.0888428 .1805059 -0.49 0.623 -.4426279 .2649424

\_IID\_14 | .3059785 .1821645 1.68 0.093 -.0510574 .6630144

\_IID\_15 | -.2261288 .2204026 -1.03 0.305 -.65811 .2058524

\_IID\_16 | -.2626337 .2182841 -1.20 0.229 -.6904627 .1651953

\_IID\_17 | .0257445 .1784158 0.14 0.885 -.3239441 .375433

\_IID\_18 | .0583754 .1717349 0.34 0.734 -.2782189 .3949697

\_IID\_19 | .1642801 .1721449 0.95 0.340 -.1731176 .5016778

\_IID\_20 | 0 (omitted)

\_IID\_21 | .0753546 .1695693 0.44 0.657 -.2569951 .4077044

\_IID\_22 | .0649273 .1751637 0.37 0.711 -.2783872 .4082419

-------------------------------------------------------------------------------

Como se puede observar, solamente las variables de alumnos por salón y el número de computadoras parecen tener efectos marginales estadísticamente significativos. En el caso del número de alumnos por salón, un aumento de 1 alumno incrementa en cerca de 1% la probabilidad de que la escuela se encuentre en una situación de alto abandono escolar, a un nivel de significancia estadística mayor al 0.01%. En el caso del número de computadoras, cada computadora adicional que tenga la escuela reduce en 1.1% la probabilidad de que la escuela experimente un alto abandono escolar, a un nivel de significancia superior a 0.01%. El efecto del número de profesores y escolaridad de los profesores son muy pequeños y no poseen un nivel de significancia estadística mayor siquiera al 10%.

En general, es posible observar que este modelo también contribuye a explicar por qué una escuela se encuentra en una situación de alto abandono escolar, controlando por los efectos individuales de cada escuela. Sin embargo, la gran mayoría de las escuelas no parecen mostrar efectos significativos indicando que esta especificación podría estar de más.

Tras haber elaborado los 4 modelos, es ahora momento de decidir cual de los modelos anteriores es más pertinente para este estudio. En el caso de los modelos de efecto fijos, estos modelos parecen mostrar dificultades por la naturaleza binaria de la variable dependiente. Por ello, la selección del modelo más adecuado para este ejercicio está entre el modelo agrupado y el modelo de efectos aleatorios. A continuación se llevará a cabo una prueba formal para determinar el modelo más adecuado.

1. Modelo agrupado contra efectos aleatorios.

Primero, se vuelve a correr el modelo logístico de efectos aleatorios:

. xtlogit abandono\_esc\_alto num\_profes escolaridad\_prof alumnos\_por\_salon num\_compu, re

Fitting comparison model:

Iteration 0: log likelihood = -168.03974

Iteration 1: log likelihood = -154.24267

Iteration 2: log likelihood = -153.9998

Iteration 3: log likelihood = -153.99969

Iteration 4: log likelihood = -153.99969

Fitting full model:

tau = 0.0 log likelihood = -153.99969

tau = 0.1 log likelihood = -151.30526

tau = 0.2 log likelihood = -149.62776

tau = 0.3 log likelihood = -148.68688

tau = 0.4 log likelihood = -148.34062

tau = 0.5 log likelihood = -148.54306

Iteration 0: log likelihood = -148.34036

Iteration 1: log likelihood = -147.58651

Iteration 2: log likelihood = -147.58375

Iteration 3: log likelihood = -147.58375

Random-effects logistic regression Number of obs = 264

Group variable: ID Number of groups = 22

Random effects u\_i ~ Gaussian Obs per group:

min = 12

avg = 12.0

max = 12

Integration method: mvaghermite Integration pts. = 12

Wald chi2(4) = 24.37

Log likelihood = -147.58375 Prob > chi2 = 0.0001

-----------------------------------------------------------------------------------

abandono\_esc\_alto | Coefficient Std. err. z P>|z| [95% conf. interval]

------------------+----------------------------------------------------------------

num\_profes | .0208071 .0174907 1.19 0.234 -.0134741 .0550883

escolaridad\_prof | -.1707759 .1091367 -1.56 0.118 -.3846799 .0431281

alumnos\_por\_salon | .0693982 .0198859 3.49 0.000 .0304226 .1083738

num\_computadoras | -.0324234 .0186547 -1.74 0.082 -.068986 .0041393

\_cons | -.1188355 1.397487 -0.09 0.932 -2.857859 2.620188

------------------+----------------------------------------------------------------

/lnsig2u | -.120745 .5609138 -1.220116 .9786259

------------------+----------------------------------------------------------------

sigma\_u | .9414138 .264026 .5433194 1.631195

rho | .2122205 .0937752 .0823405 .4471428

-----------------------------------------------------------------------------------

LR test of rho=0: chibar2(01) = 12.83 Prob >= chibar2 = 0.000

A partir de los resultados de la prueba de verosimilitud logarítmica (LR) presente en el último renglón, la probabilidad de que el modelo sin restricciones (agregado) sea mejor a un modelo constreñido (efectos aleatorios) es menor a 0.000. Por esta razón, se rechaza la hipótesis nula en favor de la hipótesis alternativa: un modelo con constreñimientos parece más adecuado para esta base de datos. Esto indicaría que, entre un modelo agrupado y otro de efectos aleatorios, el modelo de efectos aleatorios será más pertinente para esta base de datos.

1. Pruebas de cointegración y Modelo Panel Dinámico:

Puesto que la variable dependiente es binaria, no es posible llevar a cabo un Modelo de Panel Dinámico, por lo que no se llevan a cabo las pruebas de raíz unitaria o cointegración.

**IV.- Conclusiones.**

Tras llevar a cabo el ejercicio, es posible destacar algunas conclusiones. En primer lugar, se halló que, de las variables seleccionadas, solamente algunas tuvieron un efecto significativo sobre el abandono escolar. Entre éstas, el número de alumnos por salón de clase parece tener el impacto más destacado: por cada alumno adicional en el salón, aumenta en poco más de 1% la probabilidad de que la escuela experimente una situación de alto abandono escolar. Asimismo, el número de computadoras, en la mayoría de los modelos, mostró un efecto negativo sobre el abandono escolar, por lo que parecería que, por cada computadora que tenga la escuela, la deserción escolar se reduce en 0.5%. En el caso de la escolaridad promedio de los profesores por instituto, esta variable ocasionalmente salió estadísticamente significativa: un año adicional en la escolaridad promedio de los profesores podría reducir en hasta 3% la probabilidad de que la escuela presente una alta deserción escolar. El número de profesores de educación básica que tiene cada instituto no tuvo un efecto relevante o estadísticamente significativo en ninguno de los modelos.

En segundo lugar, se observó que, en términos generales, el modelo panel de efectos aleatorios fue el más pertinente para la base de datos. Al compararlo con el modelo agregado, el modelo de efectos aleatorio fue más pertinente porque asumía que existía una continuidad entre las observaciones a través de diferentes años. En el caso de los dos modelos de efectos fijos, las estimaciones de ambos modelos parecen haber sido afectadas por el comportamiento estático que ya mostraban varias de las escuelas, por lo que los efectos que indicaban parecían menos consistentes. A partir de ello, el modelo más adecuado para esta estimación fue el modelo panel estático de efectos aleatorios.

Respecto a las inferencias que se pueden hacer sobre los resultados, parecería que los datos indican que el hacinamiento y la falta de inversión en tecnología en las escuelas podrían tener un efecto importante en el abandono escolar. De esta manera, las estrategias de gobierno (en países en vías de desarrollo) para impulsar a la educación podrían ser más efectivas en reducir el abandono escolar si se enfocaran en aumentar los espacios e instalaciones educativos, al igual que invertir más en herramientas de aprendizaje valiosas para la formación y enseñanza de los alumnos. Adicionalmente, los datos parecen indicar que las estrategias de basificación masiva de docentes son insuficientes para combatir el abandono escolar, por lo que invertir en darle una mejor preparación a los futuros docentes podría tener un mayor efecto en la retención y aprendizaje de los estudiantes.

Ahora bien, estas conclusiones también tienen condiciones. Para asumir las conclusiones de este ejercicio, se tendrían que replicar estudios más recientes y en diferentes contextos, ya que el ejercicio utilizado proviene de datos de un contexto muy específico (en Brasil, entre 2003 y 2014). Asimismo, valdría la pena obtener datos adicionales para asegurase de que el efecto proviene de las variables analizadas y no sea espuria: sería útil agregar a la base de datos más información circunstancial de las escuelas, como su ubicación (i.e. una localidad pobre). Este contexto podría estar influenciando las estimaciones (i.e. los años de educación promedio de los profesores, el número de alumnos y el número de computadoras que tiene una escuela podría estar indicando el entorno económico en el que está ubicada la escuela, por lo que el efecto sobre el abandono escolar podría deberse a esa variable y no a las anteriores).

**V.- Bibliografía.**

Furger, R. (2008). How to End the Dropout Crisis: 10 Strategies for Student Retention. *George Lucas Educational Foundation, Edutopia*. <https://www.edutopia.org/student-dropout-retention-strategies>

Huberts, Alyssa y Machado, Fabiana. (2017). Brazil Education Panel Database. Banco Interamericano de Desarrollo. <https://publications.iadb.org/en/brazil-education-panel-database>

Lynch, M. (2017). How can technology be used to reduce the dropout rate? *The Tech Advocate*. <https://www.thetechedvocate.org/can-technology-used-decrease-dropout-rate/>

National Council of Teachers of English. (2014). “Why Class Size Matters Today”. <https://ncte.org/statement/why-class-size-matters/#:~:text=Overall%2C%20research%20shows%20that%20students,score%20higher%20on%20standardized%20assessments>.

Saldaña, P. (2021). “About 4 Million Students Dropped out of School during The Pandemic, Research Shows”. *Folha de S. Paulo*. Sección Internacional. <https://www1.folha.uol.com.br/internacional/en/brazil/2021/01/about-4-million-students-dropped-out-of-school-during-the-pandemic-research-shows.shtml>