

CAPÍTULO

7

Anexo Metodológico

Introducción

El *Informe Estado de la Educación* tiene como objetivo dar seguimiento al desempeño del sistema educativo costarricense. En su elaboración participa una amplia e interdisciplinaria red de investigadores, quienes utilizan diferentes técnicas e instrumentos metodológicos en sus investigaciones, con el fin de suministrar información novedosa y de calidad que sirva de insumo para propiciar el debate ciudadano en el área de la política pública educativa.

Este anexo presenta una síntesis de los principales procedimientos técnicos y metodológicos utilizados en el análisis de algunos temas incluidos en los capítulos sobre educación primaria y secundaria. Para esta edición del Informe se abordaron los siguientes tópicos: i) factores asociados a la implementación exitosa del programa de estudios de español de primero y segundo ciclos, ii) modelos geográficamente ponderados para analizar factores asociados a la repitencia en primaria y la exclusión en secundaria, iii) estimación de modelos multinivel para analizar el efecto del narcotráfico en el rendimiento académico a partir de PISA 2015, iv) modelo logit para analizar los determinantes de la aceptación de contratos docentes, y v) experimento de elección discreta para analizar los incen-

tivos para la movilidad docente a zonas vulnerables. Se espera que este apartado sirva de insumo o referencia para investigadores interesados en profundizar en el análisis del sistema educativo costarricense.

Aportes metodológicos en el capítulo de Primaria

Factores asociados a la implementación exitosa del programa de estudios de español de primero y segundo ciclos

El capítulo 3 de este Informe, sobre “Educación Primaria en Costa Rica”, documentó que los docentes de primero y segundo ciclos no cumplen con el perfil requerido por el MEP para garantizar la implementación exitosa del programa de estudios de Español en este nivel. En este sentido, resulta clave conocer cuál es la probabilidad de que el profesorado en primaria aplique con éxito la reforma curricular en esa asignatura y determinar qué factores incrementan las posibilidades de que los docentes apliquen prácticas y conocimientos adecuados para la enseñanza de los procesos de lectura y escritura.

Para aproximar estos temas, Murillo et al. (2018) estimaron un modelo de regresión logística que se muestra en la

ecuación 7.1, a partir de los datos de la encuesta Estudio sobre los procesos de lectura y escritura en la escuela primaria costarricense aplicada por el PEN para la edición de este Informe.

Primero, construyeron la variable dependiente a partir del perfil del docente “ideal” o esperado para la enseñanza de la lectura y escritura, es decir, aquel que posee los conocimientos y prácticas adecuadas para la implementación exitosa del programa de estudios de español. Para ello se tomó como referencia una selección de ítems de la encuesta que pueden ser consultados en el capítulo 3 de este Informe.

Segundo, en el caso de las variables independientes, se crearon tres índices: utilización de recursos en el aula, factores que favorecen la implementación, competencias desarrolladas durante la formación universitaria para los procesos de lectura y escritura. En este procedimiento se aplicó un análisis factorial para identificar cuáles ítems de la encuesta componían cada índice, luego, se estandarizaron en la estimación final del modelo. La composición de los índices, así como los coeficientes, razones de probabilidad (odds ratio) y errores estándar están disponibles en la ponencia Murillo et al. (2018), en el sitio www.estadonacion.or.cr.

Ecuación 7.1

Modelo de factores asociados a la probabilidad de que los docentes empleen las prácticas de enseñanza de lectura y escritura según la reforma curricular de Español en primero y segundo ciclos

$$P(PE) = \beta_0 + \beta_1 \text{Edad} + \beta_2 \text{Sexo} + \beta_3 \text{CompetenciasU} + \beta_4 \text{RecursosAula} + \beta_5 \text{EnfoqueTradi} + \varepsilon_i$$

Donde:

P(PE): es la probabilidad de emplear prácticas de enseñanza de lectura y escritura según la reforma curricular de Español.

Edad: es una variable continua de la edad en años.

Sexo: es una variable dicotómica para el sexo del docente (1: mujer, 0: hombre).

CompetenciasU: variable continua del índice de competencias universitarias de los docentes encuestados

RecursosAula: variable continua del índice de uso de recursos en el aula por parte de los docentes encuestados

EnfoqueTradi: variable continua del índice de empleo de prácticas de enseñanza bajo el enfoque tradicional, por parte de los docentes encuestados.

ε_i : es el término de error del modelo.

Ecuación 7.2

El estadístico de Moran

$$I = \left[\frac{N}{S} \right] \cdot \{(e'W e) | e'e\}$$

Donde:

e es un vector de residuos del modelo.

N es el número de unidades espaciales indexadas.

S es un factor de estandarización igual a la suma de todos los elementos de la matriz de distancias (*W*).

PARA MÁS INFORMACIÓN SOBRE CONCEPCIONES DE LOS DOCENTES ACERCA DE LA ENSEÑANZA DE LA LECTURA Y LA ESCRITURA

véase Murillo, et al., 2018 en www.estadonacion.or.cr

Modelos geográficamente ponderados para analizar factores asociados a la repitencia en primaria

Un estudio realizado, para esta edición del Informe, analizó a los centros educativos de primaria con mayor exposición al tráfico de drogas, a través de la estimación de modelos econométricos geográficamente ponderados (GWR por sus siglas en inglés, Fotheringham et al., 2002). El uso de GWR es una técnica de regresión espacial que proporciona un modelo local de la variable o proceso que intenta entender/prever al ajustar una ecuación de regresión a cada entidad en el conjunto de datos.

Sánchez (2019) analizó la situación de repitencia en las escuelas que registraban más de 1.500 eventos al año, asociados al narcotráfico, en un radio menor a 1.000 metros. Las estimaciones realizadas incluyeron un registro de 529

Cuadro 7.1

Estimación de los estadísticos de Moran y los multiplicadores de Lagrange para el análisis de repitencia en primaria

Prueba	Estadístico	Valor de P ^{a/}
Estadístico de Moran	43,2	0,000
<i>Error espacial</i>		
Multiplicador de Lagrange	289,2	0,000
Multiplicador robusto de Lagrange	2,0	0,051
<i>Rezago espacial</i>		
Multiplicador de Lagrange	551,0	0,000
Multiplicador robusto de Lagrange	123,2	0,000

a/ Se refiere al valor de significancia estadística.

Fuente: Sánchez, 2019.

centros educativos de primaria. Dentro de las variables explicativas se incluyó el nivel de pobreza (medido a través del método de necesidades básicas insatisfechas, NBI), la tasa de incautaciones de drogas por 100 habitantes, la tasa de homicidios por 100 habitantes, la tasa de desempleo, el bajo clima educativo, el porcentaje de profesores interinos, el porcentaje de profesores en propiedad, el porcentaje de profesores aspirantes, la relación alumnos/profesor, el tamaño del

colegio (matricula), la edad promedio de los docentes y finalmente, un indicador que agrupa la inversión social construido a partir de la inversión por estudiante promedio en becas, alimentación, transporte e infraestructura.

Para analizar la existencia de correlación espacial en los modelos planteados, se siguieron dos pasos. Primero se corrió un modelo con las variables indicadas y luego se calculó el estadístico de Moran (ecuación 7.2 y cuadro 7.1).

Finalmente, se estimó el modelo utilizando el método de máxima verosimilitud, los resultados se muestran en el cuadro 7.2. Como se puede observar las variables tasa de incautaciones de drogas, subsidio en transporte y estudiantes por docente dieron significativas y muestran una relación espacial con el comportamiento que presenta la repitencia en las escuelas. Un mayor detalle de los resultados de este modelo se puede consultar en el estudio de Sánchez (2019), disponible en el sitio www.estadonacion.or.cr.

PARA MÁS INFORMACIÓN SOBRE PATRONES ESPACIALES DEL TRÁFICO DE DROGAS

véase Sánchez, 2019 en www.estadonacion.or.cr

Aportes metodológicos en el capítulo de Secundaria

Modelos geográficamente ponderados para analizar factores relacionados a la exclusión educativa

Para analizar los efectos del narcotráfico y otro conjunto de variables sociodemográficas sobre la exclusión educativa en secundaria, Sánchez (2019) utilizó regresiones ponderadas geográficamente (GWR). En el análisis se utilizó como variable dependiente la tasa de exclusión intraanual en colegios públicos para el año 2017. Por su parte, las variables independientes fueron agrupadas en tres dimensiones: el contexto en el cual se ubica el colegio, las características de los docentes y del centro educativo, y la inversión en programas sociales en educación (cuadro 7.3).

Con la información anterior, Sánchez (2019) desarrolló modelos globales (para todo el país) y modelos locales (por área de influencia distrital). En el primer caso, se parte del supuesto de que los factores analizados influyen de manera similar en todos los colegios², por tanto, los parámetros estimados se interpretan como un valor promedio a nivel nacional.

Cuadro 7.2

Resultados del modelo global geográficamente ponderado para analizar la repitencia en primaria^{a/}

Variable	Coefficiente	Estadístico	Valor P ^{b/}	
Constante	0,229	8,902	0,0001	***
Tasa de incautaciones	0,864	0,219	0,0144	**
Subsidio de transporte	-0,981	0,991	0,0431	**
Relación estudiantes/docentes	0,472	2,875	0,0514	*

a/ El modelo fue estimado con un total de 529 observaciones, y con un valor de R² igual a 0,562.

b/ Los resultados solo muestran las variables estadísticamente significativas al 99% (***) , 95% (**) y 90% (*).

Fuente: Sánchez, 2019.

Cuadro 7.3

Variables utilizadas en el modelo geográficamente ponderado para analizar la exclusión en colegios públicos

Dimensión	Variables
Contexto en el cual se ubica el colegio	<ul style="list-style-type: none"> • Pobreza, medido a través del método de necesidades básicas insatisfechas (NBI). • Tasa de incautaciones de drogas por 100 habitantes. • Tasa de homicidios por 100 habitantes. • Tasa de desempleo. • Bajo clima educativo.
Características de los docentes y el centro educativo	<ul style="list-style-type: none"> • Porcentaje de profesores interinos, en propiedad y aspirantes. • La relación alumnos/profesor. • El tamaño del colegio, medido a través de la matrícula.
Inversión en programas sociales en educación.	Inversión promedio por estudiante en: <ul style="list-style-type: none"> • Becas. • Alimentación. • Transporte. • Infraestructura.

Fuente: Sánchez, 2019.

En el segundo caso, se construye una ecuación diferente para cada área de influencia distrital mediante la incorporación de la variable dependiente e independientes para cada caso. Según Fotheringham et al. (2002) las GWR permiten estimar parámetros locales, ya que se reconoce que pueden existir variaciones espaciales en las relaciones anali-

zadas, de modo que existen coeficientes con valores distintos para cada área.

Para analizar la existencia de correlación espacial en los modelos planteados, se siguieron dos pasos. Primero se corrió un modelo con las variables indicadas en el cuadro 7.3 y luego se calculó el estadístico de Moran (ecuación 7.2). Al estimar el modelo inicial se descartó un

conjunto de variables que presentaban bajos niveles de significancia, heterocedasticidad³ y multicolinealidad⁴.

Posteriormente, se aplicaron las pruebas de multiplicadores de Lagrange para determinar cuál modelo (de rezago o de error) era mejor para la estimación (Sánchez, 2019). Los resultados obtenidos se presentan en el cuadro 7.4, el multiplicador robusto de Lagrange⁵ evidencia que el mejor modelo es el de rezago espacial, ya que resultó significativo.

Los resultados generales del modelo global geográficamente ponderado para analizar la exclusión intraanual en los colegios públicos, adicionando un término de rezago espacial, se presentan en el cuadro 7.5. Sánchez (2019), además, utilizó el método de máxima verosimilitud debido a la presencia de correlación espacial.

PARA MÁS INFORMACIÓN SOBRE PATRONES ESPACIALES DEL TRÁFICO DE DROGAS

véase Sánchez, 2019 en www.estadonacion.or.cr

Modelos multinivel para explorar la relación entre los resultados en PISA 2015 y narcotráfico

Ediciones anteriores de este Informe dieron seguimiento a los resultados obtenidos por los estudiantes del sistema educativo costarricenses en las pruebas PISA 2012 y PISA 2015, pero no se profundizó en el efecto de la violencia criminal sobre los resultados académicos en esas pruebas. En esta edición se realiza un primer estudio exploratorio en el cual Barrado (2018) utiliza la metodología Modelos Multinivel para estimar si la exposición al tráfico de drogas — medida a través de los kilogramos de cocaína incautados⁶— afecta significativamente al rendimiento académico de los estudiantes en esas pruebas. El modelo base puede expresarse como una función clásica de producción educativa (ecuación 7.3) propuesta por Hanushek et al. (2013).

Cuadro 7.4

Estadístico de Moran y multiplicadores de Lagrange para el análisis de exclusión en secundaria

Prueba	Estadístico	Valor de P ^{a/}
Estadístico de Moran	25,0	0,000
<i>Error espacial</i>		
Multiplicador de Lagrange	289,2	0,000
Multiplicador robusto de Lagrange	0,7	0,311
<i>Rezago espacial</i>		
Multiplicador de Lagrange	422,6	0,000
Multiplicador robusto de Lagrange	71,1	0,000

a/ Se refiere al valor de significancia estadística.

Fuente: Sánchez, 2019.

Cuadro 7.5

Resultados^{a/} del modelo global geográficamente ponderado para analizar la exclusión en secundaria

Variable	Coefficiente	Estadístico	Valor P ^{b/}
Constante	0,0798	5,420	0,00001 ***
Tamaño	-0,1996	-1,545	0,00001 ***
Relación estudiantes/docentes	0,2541	2,471	0,00001 ***
Tasa de incautaciones	0,9515	0,344	0,05110 *
Pobreza	0,6134	1,183	0,00152 ***
Tasa de desempleo	0,4058	1,722	0,03552 **
Subsidio de alimentos	-1,7236	-1,429	0,04386 **
Porcentaje de profesores interinos	-0,1707	2,082	0,00001 ***

a/ El modelo fue estimado con un total de 723 observaciones, el ancho de banda es igual a 16.211,97 y el valor de R² igual a 0,5808.

b/ Los resultados solo muestran las variables estadísticamente significativas al 99% (***) 95% (**) y 90% (*).

Fuente: Sánchez, 2019.

Dentro del vector de características individuales se incluyó variables como género, edad y país de nacimiento. En el caso de las familiares, la educación de los padres, su ocupación y la cantidad de libros en el hogar, como *proxy*, del nivel socioeconómico y cultural. En relación con el centro educativo, el tamaño de la escuela, la titularidad, la ubicación de la escuela, la proporción de profesores certificados y el nivel de autonomía del centro. Adicionalmente, se incluye la situación socioeconómica del distrito

geográfico donde se sitúa el centro escolar como variable de control. En este análisis, el narcotráfico se mide a través de la tasa de kilogramos de cocaína incautados por cada 100.000 habitantes a nivel cantonal.

De acuerdo con Barrado (2018) la literatura existente sobre las consecuencias del narcotráfico y la violencia han mostrado que escuelas situadas en vecindarios inseguros, con altos niveles de violencia o presencia de grupos involucrados en el narcotráfico presentan mayores niveles

Ecuación 7.3

Función de producción educativa

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 N_i + \varepsilon_i$$

Donde:

Y_i denota el puntaje medio de los cinco valores plausibles del alumno i -ésimo en las pruebas de matemáticas, lectura y ciencias de PISA 2015.

X_i es un vector de características individuales y familiares.

Z_i representa al vector de características de la escuela y del distrito.

N_i representa la exposición al narcotráfico a nivel cantonal.

ε_i es el término de error propio de cada alumno, que recoge los factores no observados y los errores de medición.

Ecuación 7.4

Función para estimar la probabilidad de que un docente de secundaria acepte un contrato en el MEP

$$P(\text{Aceptación del contrato}_i = 1 | x) = \beta_0 + \beta_1 \text{Docente}_j + \beta_2 \text{Colegio}_k + \beta_3 \text{Distrito}_l + EF_{\text{Cantón}} + \varepsilon_i$$

Donde:

$Y_i P(\text{Aceptación del contrato}_i = 1 | x)$ es la probabilidad de aceptación de un contrato i .

Docente, se refiere a un vector con las características del docente.

Colegio, representa el vector con las características del centro educativo de destino y sus estudiantes

Distrito, es un vector con las características del distrito donde se encuentra localizado el colegio en el cual se incluye nuestra variable principal: la pobreza.

En la ecuación, además se incluyen efectos fijos por cantón $EF_{\text{Cantón}}$ para controlar por la presencia de heterogeneidad no observada entre cantones y comparar los contratos ofrecidos en el mismo cantón.

ε_i es el término de error en el modelo.

Modelo logit para analizar los determinantes de la aceptación de contratos docentes

Un estudio exploratorio realizado por Lentini (2019b) estimó la probabilidad de que un docente acepte un contrato en el MEP. Para ello se utilizó una base de datos creada a partir de dos fuentes: la primera, son los registros administrativos del Ministerio de Educación Pública (MEP) con información sobre todos los contratos aceptados y rechazados por los docentes de educación secundaria pública del país y sobre las características de los colegios donde se ofrecen los contratos; y la segunda, del Censo de Población y Vivienda de 2011 del INEC con información sobre la población en los distritos donde se localizan los centros educativos (proyectada a 2017). El archivo de datos quedó constituido por 6.070 contratos, con información de 4.484 docentes, en 650 colegios y 339 distritos⁹.

Se empleó una regresión logit con la aceptación de los contratos como variable dependiente. La estrategia para identificar los determinantes de la aceptación de contratos por parte de los docentes de Costa Rica fue calcular su probabilidad en dependencia de factores en tres niveles de análisis, las características de: a) los docentes a quienes se les ofrece el contrato, b) los centros educativos, y c) la localización de los colegios. En la ecuación 7.4 presenta una extensión del modelo propuesto por Barbieri et al. (2011) utilizado en la estimación.

Las variables explicativas incluidas en cada uno de los vectores fueron las siguientes: a) características de los docentes a quienes se les ofrecía el contrato: el género, la edad, la edad al cuadrado, la provincia de nacimiento del docente, la coincidencia geográfica (si la provincia donde se ofrecía el contrato coincidía con la de nacimiento del docente), la categoría profesional, si la asignatura que debía impartir era científica (Matemáticas, Biología, Química, Física, Informática), y si la asignatura a impartir era básica (Matemáticas, Estudios Sociales, Español, Inglés, y alguna ciencia – Biología, Química, Física); b) características de los colegios de destino: porcentaje de estudiantes suspendidos, relación

o tasas de absentismo escolar, lo que puede deteriorar el rendimiento académico (Romano, 2015; Jarillo et al., 2016). Con el objetivo de controlar esa situación se incluyó un indicador que recoge la frecuencia de absentismo estudiantil.

Un problema común en el análisis de datos de corte transversal es que existen un conjunto de características no observables que pueden afectar al rendimiento escolar y que no están incluidas en el modelo. Para considerar esta cuestión se recurre al uso de Modelos Jerárquico Lineales, también denominados Modelos Multinivel.

Específicamente, se trabajó con dos niveles: estudiantes y centros escolares. Adicionalmente, al objeto de evitar una mala interpretación del poder explicativo de las variables del modelo, se incluyen errores estándar robustos mediante la metodología clúster, usando cada distrito como variable clúster o grupo de corrección. Este método contempla la existencia de variables explicativas o características no observables, inherentes a cada uno de los distritos y no recogidas por el modelo.

estudiantes/docentes, tipo de colegio (académico o técnico) y cantidad de horas que incluía el contrato; y c) características de los distritos de destino: porcentaje de población en pobreza, población extranjera, porcentaje de población activa en el mercado laboral, porcentaje de población de personas que trabajan en el sector terciario, importancia del distrito en el cantón, cantidad de colegios en el distrito, y el tamaño del cantón al que pertenecía el distrito, y clasificación del MEP por zonas (urbano-rural).

Los resultados se presentaron mediante razones de probabilidad (ecuación 7.5) para los estimadores. Este es el cociente entre la probabilidad de que el contrato sea aceptado y la probabilidad de que no lo sea.

Experimento de elección discreta (EED) para analizar los incentivos para la movilidad docente a zonas vulnerables

Otro estudio, realizado por Lentini (2019a), identificó las preferencias de los docentes para aceptar un contrato temporal en centros poco atractivos fuera del GAM. Para ello utilizó un experimento de elección discreta (EED) que es una técnica cuantitativa para la identificación de preferencias de los entrevistados ante escenarios hipotéticos alternativos que se describen en términos de atributos y niveles. La importancia de los atributos en la aceptación de los escenarios se estima a través de modelos de utilidad aleatoria (McFadden, 1973). Esta investigación con docentes de Costa Rica se llevó a cabo mediante un EED diseñado en dos fases: una cualitativa y otra cuantitativa.

La fase cualitativa siguió un proceso en varias etapas que inició con la revisión de literatura, y continuó con entrevistas profundas a informantes clave y docentes del perfil meta. Las consultas se llevaron a cabo siguiendo una guía de entrevista semi-estructurada basada en Rao et al. (2010). En esta fase se identificó una lista de potenciales atributos, la cual se restringió a seis, para evitar la complejidad de las alternativas y de la selección de los contratos¹¹.

De los seis atributos seleccionados que se valoraban en la fase cuantitativa, dos

Ecuación 7.5

Razón de probabilidad (*odds ratio*) en la probabilidad de aceptar un contrato

$$\text{odds ratio } (P) = \frac{P}{1-P}$$

Donde:

P es la probabilidad de aceptación del contrato.

Las razones de probabilidad pueden tomar un valor de 0 a infinito. Se interpreta que la variable aumenta la posibilidad de ocurrencia del evento si el odds ratio es mayor a 1, y que la reduce si es menor. Se calcula los odds ratio en función de intervalos de confianza para valorar su significatividad.

estaban relacionados directamente con el ejercicio de la docencia: i) equipo profesional de apoyo con acompañamiento pedagógico; ii) recursos tecnológicos y materiales pedagógicos; y los otros cuatro eran más de beneficio individual como: iii) ingreso adicional como porcentaje de su salario base mensual actual por el periodo de traslado temporal (30%, 40% y 50% sobre el salario base); iv) puntaje para ser propietario o categoría profesional docente; y v) vivienda prestada para la familia (Lentini, 2019a).

Una vez seleccionados los atributos y niveles, se procedió a diseñar una primera versión de cuestionario para ser revisado en sesiones grupales con nuevos informantes clave y fue probado con informantes meta (Lancsar y Louviere, 2008; Klotjgaard et al., 2012). Los escenarios del EEC fueron definidos utilizando el Software Sawtooth Choice-Based Conjoint (Sawtooth Software Inc. 2007) y un diseño factorial fraccional.

La información de la fase cuantitativa se recolectó mediante una encuesta cara a cara a 400 docentes de asignaturas básicas de una muestra representativa de 52 colegios públicos. Estos fueron seleccionados aleatoriamente mediante un proceso multietápico estratificado (por regiones), con afijación proporcional por cantidad de docentes y por lo tanto es representativa de la región central. Los datos se recolectaron del 18 de febrero

al 30 de marzo de 2019 por la empresa Unimer¹².

El procesamiento de los datos se llevó a cabo con el programa Stata mediante un modelo *logit mixto* que ha sido generalmente utilizado para el análisis de EED (Li et al., 2014). Dentro de este marco, se asume que el individuo *n* es racional y se enfrenta a una selección entre *C* alternativas de contrato (en nuestro caso en cada escenario tiene dos alternativas de contratos posibles *C=2*) y opta por aquella que le brinda un mayor beneficio o satisfacción. Por lo tanto, el individuo *n* escoge la alternativa *i* sobre la *j* si y solo si: $U_{ni} > U_{nj}$. Donde *U* es la utilidad para una alternativa específica (ecuación 7.6).

Lentini (2019a) utilizó el método de *logit mixto* (MXL), el cual emplea técnicas de estimación basadas en la simulación para lo cual se definen una cantidad de sorteos Halton que indica el número veces el MXL hará simulaciones¹³. El MXL tiene la particularidad de que permite que se viole el supuesto de la independencia de alternativas (Vuijicic et al., 2010). Esta posibilidad hace que se puedan obtener múltiples observaciones por individuo en la estimación de los parámetros, y que estos varíen de acuerdo con su propia distribución estadística. Adicionalmente, con el MXL las heterogeneidades de las preferencias de la muestra quedan incorporadas en el modelo porque los coeficientes son trata-

Ecuación 7.6

Función de utilidad y modelo logit de selección para la escogencia de una alternativa específica

La utilidad del individuo n que escoge la alternativa i (U_{ni}) no es directamente observable y está dada por:

$$U_{ni} = V_{ni} + \varepsilon_{ni}$$

Donde:

P es la probabilidad de aceptación del contrato.

V_{ni} representa el componente determinístico, es una función de los atributos del contrato (x_1, \dots, x_m) que pueden ser observados (bono salarial, vivienda, transporte, entre otros).

ε_{ni} es el componente aleatorio, conformado por los atributos no observables del contrato y de las variaciones de las preferencias individuales de las personas. Se supone que ε_{ni} están independientes e idénticamente distribuidos (IID)

En un modelo *logit* de selección, la probabilidad de escoger la alternativa i sobre la j , está dado por:

$$P_{ni} = Pr [U_{ni} > U_{nj}] = Pr [V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}] = Pr [\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj}]$$

Tal que la utilidad de la persona n se obtiene de calcular la utilidad de seleccionar alternativa i sobre la alternativa j . La especificación del modelo puede reescribirse de la siguiente forma:

$$U_{ni} = V_{ni} + \varepsilon_{ni} = \alpha_i + \beta_1 x_{1ni} + \beta_2 x_{2ni} + \dots + \beta_m x_{mni} + \varepsilon_{ni}$$

Donde β , son los coeficientes de los atributos de cada alternativa. Todos los atributos y sus niveles son codificados como variables dicotómicas. El nivel más bajo de cada atributo queda fuera de la regresión.

PARA MÁS INFORMACIÓN SOBRE

FACTORES ASOCIADOS A LAS PREFERENCIAS LABORALES DE LOS DOCENTES

véase Lentini, 2019a en www.estadonacion.or.cr

dos como aleatorios y no como si fueran fijos (Flynn et al., 2010; Kruk et al. 2010; Fiebig et al., 2011).

Se hicieron dos análisis, uno con la muestra total de 400 docentes (forzada), y otro solo con quienes escogían alguna alternativa de los contratos. La segunda

muestra se redujo a 209 docentes, ya que se eliminan los casos en los que los docentes no escogieron ninguna alternativa. Es decir, los docentes que bajo ninguna de las posibilidades de contrato que se ofrecían aceptaban el desplazamiento.

Este Anexo Metodológico fue preparado por:

Dagoberto Murillo, Katherine Barquero y Valeria Lentini.

Notas

1 Estos índices obtuvieron una confiabilidad estadística superior al 70%.

2 De acuerdo con Sánchez (2019) este es un supuesto fuerte dadas las diferencias territoriales existentes.

3 En Estadística se dice que existe heterocedasticidad cuando los errores no son constantes a lo largo de toda la muestra.

4 En Estadística la multicolinealidad ocurre cuando algunas variables predictoras incluidas en el modelo están correlacionadas con otras variables predictoras.

5 Este tipo de prueba se aplica cuando los multiplicadores de Lagrange para las dos especificaciones, error espacial o rezago espacial, son significativos.

6 De acuerdo con Barrado (2018) los kilogramos incautados constituyen el indicador más utilizado para medir las actividades de tráfico a nivel regional.

7 Ya sea a nivel individual, escolar y cantonal.

8 Esta técnica permite obtener estimadores insesgados ya que se adapta a la naturaleza jerárquica de los datos evitando así los problemas derivados de la presencia de autocorrelación. Además, los modelos multinivel están compuestos por dos partes diferenciadas, una común y fija para todos los contextos y una segunda parte que varía y se estima en función de cada nivel (Barrado, 2018).

9 Se eliminaron aquellos contratos de menos de 20 horas (menos de tiempo completo), para evitar que los resultados estuvieran sesgados por esta variable ya que los contratos por pocas horas suelen ser complementados por los docentes con contratos adicionales.

Las estrategias de los docentes para completar esas horas—dentro o fuera del sistema— que incidieran en su decisión de aceptar o no los contratos no serían posibles de observar ni controlar mediante el modelo. También se eliminaron aquellos ofrecidos a docentes con contratos en propiedad (Lentini, 2019b).

10 Los informantes clave incluyeron un exministro de Educación, funcionarios del departamento de contratación y recursos humanos del Ministerio de Educación, representantes gremiales, asociación de profesionales docentes, abogados especialistas en políticas de contratación docente e investigadores académicos especialistas en Educación.

11 Para esta selección se siguieron dos criterios: primero, la frecuencia con la cual el atributo había sido mencionado en las entrevistas con docentes, y el nivel de prioridad que estos daban al atributo en la decisión de aceptar una oferta de contrato, y segundo, la valoración que hacían los informantes clave sobre la viabilidad de ofrecer el atributo (Lentini, 2019a).

12 Los docentes no fueron compensados por hacer la entrevista, es decir, fue voluntaria. Se hacía un máximo de 8 entrevistas por centro educativo (2 docentes máximo por materia) para evitar la aglomeración de observaciones. La encuesta se realizaba entre lecciones o en el tiempo libre de los docentes. La duración promedio de cada entrevista fue de 25 minutos.

13 Es importante correr una cantidad lo suficientemente grande para que el modelo converja y que los estimadores se estabilicen. Es decir, que la varianza inducida en la simulación se minimice (WHO, 2012). El MXL tiene la particularidad de que permite que se viole el supuesto de la independencia de alternativas (Vuijic et al., 2010).