

ANÁLISIS URBANO Y MODELOS MULTINIVEL. UNA INTRODUCCIÓN

Henio Millán Valenzuela. Profesor Investigador de El Colegio Mexiquense hmillan@cmq.edu.mx

Fechas del Artículo:
Recibido: 22/01/2020
Dictamen: 18/03/2020
Aceptado: 11/02/2020
Publicado: 30/05/2020
Orcid: 0000-0003-0115-0636

Forma de citar este artículo: Millán H. (2020). Análisis urbano y modelos multinivel. Una introducción, *Revista Científica de Estudios Urbano Regionales Hatsö-Hnini*, Año 1 Vol 1 No. 2 -pp. 3-21 disponible en www.revistahatsohinini.com.mx México



Resumen. El objetivo es mostrar cómo pueden usarse los modelos multinivel para el análisis urbano. Se utiliza un ejemplo con datos ficticios y se exponen varios modelos para captar diferencias en los precios barriales, a través de medias y de pendientes aleatorias. Estos precios son la variable dependiente, mientras que las independientes son el ingreso, el índice hedónico y el mercado de tierras. Las tres primeras se registran al nivel más elemental de las colonias, y la última, en el segundo, que corresponde a las ciudades. La conclusión es que este tipo de ejercicios pueden ser una herramienta valiosa a la hora de explicar discrepancias inter e intracitadinas de precios barriales

Palabras clave: *Análisis multinivel. Herramienta. Precios prediales. Colonias. Ciudades*

Abstract. The goal is to show how multilevel models can be used for urban analysis. An example with fictional data is used, and various multilevel models are exposed to capture differences in quarter prices of land, through random means and slopes. These prices are the dependent variable, while mean income, hedonic index and land market are the independent ones. The first three are registered at quarter-first level; the last one, on the city-second level. Main conclusion is that this kind of exercise may be a valuable tool when we try to explain inter and intra-cities quarter prices

Key word. *Multilevel Analysis. Tool. Land Prices. Quarters. Cities*

Introducción

El uso de los modelos multinivel o jerárquicos tiende a proliferar en el campo de las ciencias sociales. La razón es la inquietud por la diversidad. Ahí donde las unidades más elementales pueden ser

aglutinadas en grupos que comparten alguna o varias características en común, surgen preguntas en torno a diferencias que ocultan las proposiciones generales, activadas por los métodos cuantitativos tradicionales. Estas inferencias suelen ser válidas para el conjunto de la población bajo estudio, pero por la naturaleza de los procedimientos estadísticos pueden ocultar la heterogeneidad que subyace tanto entre sus componentes individuales como entre los propios grupos que los conglomeran. Atestiguar y explicar esas disparidades es el principal problema que pretenden encarar los modelos multinivel.

Para muestra bastan dos botones: como todos sabemos, en una sociedad extremadamente polarizada, con muy pocos ricos y muchos pobres, la media de ingreso encierra discrepancias tan grandes, que resulta irrelevante cuando queremos forjarnos una imagen de su bienestar. Las medidas de dispersión nos advierten la existencia de tales diferencias, así como de la cautela que debemos tener al utilizar este indicador. Pero no nos explican la causa de estas diferencias. Por otro lado, podemos aventurar un modelo de regresión que trate dar cuenta de estas diferencias mediante una ecuación en la que el nivel de ingreso dependa del capital humano por ejemplo, el nivel de educación de los individuos. El coeficiente estimado del regresor nos dirá que por cada unidad de aquel capital que distingue a un individuo de otro, el ingreso de ambos discrepará en una determinada proporción *fija*, suministrada por el valor del coeficiente de la regresión. El carácter fijo de semejante estimación quiere decir: general. El mensaje, entonces, es inequívoco: dos personas con el mismo capital humano deberán de gozar, *en promedio*, de un ingreso similar. De acuerdo al modelo lineal clásico, cualquier variación en el ingreso para un mismo nivel de educación sólo puede obedecer a:

i) los residuos o errores, que miden la distancia entre el pronóstico y el dato efectivo, que constituyen la parte no explicada por el modelo; ii) la falta de inclusión de otras variables explicativas (independientes), que pueden aumentar la proporción explicada por el mismo, si resultan pertinentes.

Sin embargo, hay otra salida que el análisis de regresión no toma en cuenta: los grupos en los que puede dividirse la población muestral bajo escrutinio. Supongamos que los que lograron egresar de una universidad son aglutinados por la licenciatura que escogieron, y escojamos dos conjuntos extremos: los teólogos y los financieros. Es razonable pensar que, a pesar del mismo nivel de estudios, los ingresos de ambos grupos sean distintos. Ello significa que el monto de las remuneraciones laborales que responden a cada año de estudio varía de un grupo a otro. De otra forma: el coeficiente del regresor que da cuenta de esa respuesta no sólo puede variar; además, esta variación representa una versión más fiel de la realidad y un campo de estudio adicional.

Estos dos ejemplos ilustran el foco de atención de los modelos multinivel (ML): las diferencias que prevalecen entre las medias de los grupos, por un lado, y entre las pendientes de esos conglomerados, por el otro. Sin embargo, no soslayan otra de las dimensiones de la diversidad: la diferencia intergrupala. De esta forma, se completa el análisis de las diferencias: las que se desarrollan *dentro* de los grupos y las que se registran entre uno y otro conjunto.

El lector advertirá que la agrupación es el dispositivo que permite activar el análisis multinivel; motivo suficiente para ofrecerse como una herramienta atractiva para las ciencias sociales empíricamente sustentadas. Al estudiar individuos que despliegan su vida al amparo de alguna modalidad societal, pero lo efectúan de forma segmentada, el antiguo interés por los aspectos comunes abre paso al examen de las diferencias. Y, por esta vía, al de las responsabilidades que pueden ser atribuidas a cada uno de los niveles en los que se edifica una comunidad: el individual, el grupal y el macrosocial.

Este es el caso de los fenómenos urbanos: los individuos se aglutinan en viviendas; y éstas en colonias o barrios, que a su vez conforman un conglomerado ciudadano. Aunque las casahabitación de un mismo barrio suelen parecerse,

también presentan rasgos distintos; las diferencias entre barrios pueden ser notables, aún cuando sean colindantes, si la ciudad atestigua disparidades sensibles en el nivel de bienestar. Por tal razón, el multinivel es un instrumento que puede aplicarse provechosamente a la hora de estudiar aquellos fenómenos.

Este artículo persigue mostrar el instrumental básico de los modelos multinivel, con el propósito de que los estudiosos de las dinámicas urbanas se acerquen a sus procedimientos y técnicas. *Su preocupación central es, entonces, de índole pedagógica; no, investigativa.* Por tal razón, *los datos y resultados cuantitativos son hipotéticos, inventados si se quiere, y no el fruto de un examen genuinamente empírico.* El ejemplo adoptado explora las diferencias en el precio promedio del metro cuadrado de los terrenos para casahabitación en un conjunto de ciudades, que aglutinan colonias o barrios y forman parte de un territorio nacional.

Este ejercicio se organiza en seis secciones. La primera articula la introducción de las variables y su ubicación por niveles con una pregunta de investigación y con la hipótesis principal. La segunda se orienta a mostrar los procedimientos que avalan si las diferencias entre precios barriales son o no estadísticamente significativas; la segunda introduce una variable independiente orientada a explicar tales diferencias; la tercera recurre a una variable de control, con la intención de aquilatar hasta qué punto la variación interbarrial puede ser atribuida exclusivamente al principal determinante, y hasta dónde está condicionada por la influencia de otros factores.

El cuarto y el quinto apartados abordan el tema de las pendientes; es decir, de la respuesta de los precios de los terrenos a las diferencias en la variable independiente protagonista. En uno, nos preguntamos si tal reacción acusa registros distintos entre los barrios, mientras en el siguiente se inspecciona los factores de estas discrepancias.

I. Variables y ubicación por niveles.

1. El objeto de estudio y la hipótesis como determinantes de las variables y de los niveles.

Si la agrupación es el detonante de los modelos multinivel, el primer paso consiste en especificar los distintos niveles en los que se ubican las unidades más elementales y las instancias que los

agrupan. Si llamamos “precio barrial” al promedio de los precios del metro cuadrado en una colonia o barrio, la pregunta de investigación puede formularse de la siguiente manera: ¿por qué los precios barriales varían tanto entre las colonias como entre las ciudades de un mismo territorio urbano nacional?

Esta cuestión imprime a las colonias o barrios la categoría de unidad elemental. Por lo tanto, ella y sus atributos corresponden al nivel 1. Las ciudades, en cambio, conforman los grupos que las aglutinan; sus atributos, expresados en variables explicativas, deben ser ubicadas en el nivel 2. Lo resaltante es que es la pregunta de investigación y, específicamente, el objeto de estudio, el que determina el nivel. Este precio funge como variable dependiente y siempre se ubica en el nivel 1.

Las hipótesis adelantan respuestas a la pregunta central del estudio. Cuando se transforman en proposiciones operativas a partir de juicios conceptuales, se convierten en relaciones causales entre la variable que queremos explicar y las variables explicativas o independientes. Por ejemplo, si nuestra primera hipótesis es que las diferencias entre los precios barriales obedecen fundamentalmente al estatus socioeconómico de las colonias, éste deviene la principal variable independiente. Sin embargo, aquí enfrentamos dos problemas: primero, ¿cómo medir el estatus económico? Si optamos por una clasificación estratificada (alto, mediano, bajo), tendríamos una variable nominal categórica, que es perfectamente válida tanto para el análisis de regresión como para el de multinivel. ¿Pero cómo dibujamos los umbrales entre un estrato u otro?. El método Dalenius permite la estratificación de una variable continua, pero si se cuenta con ella, es más conveniente recurrir directamente a ella. Por tal razón, escogemos el promedio barrial del ingreso familiar como *proxy* del estatus socioeconómico.

El segundo problema es en cuál nivel se ubica la principal variable independiente. El ingreso medio barrial es un atributo de la colonia, no de la ciudad que la alberga. Por tanto, esta variable debe ubicarse en el nivel 1. Entonces, una vez que se torna operativa, la hipótesis principal puede expresarse de la siguiente forma:

$$P_{ij} = \beta_{0i} + \beta_{1i}X_{ij} + e_{ij}$$

Donde P_{ij} es el precio barrial de los terrenos de la colonia i -ésima, ubicada en la ciudad j -ésima. β_{0i} es el intercepto que corresponde a la ecuación de regresión que vincula este precio con el ingreso medio

barrial (X). β_{1i} , es la pendiente esta ecuación e indica cuánto varía el precio cuando X cambia en una unidad (un peso, un dólar, etc.). Por último, e_{ij} es el residuo u error que emerge de las discrepancias entre el valor efectivo del precio de la colonia i -ésima, ubicada en la ciudad j -ésima, y el estimado por el modelo. Corresponde, por tanto, a la porción no explicada por este ejercicio econométrico.

Hasta aquí se ha formulado en lenguaje operativo la hipótesis central de la investigación. Sin embargo, ésta puede aceptar matices o condicionantes. Si el examen minucioso de los datos nos arroja una imagen en la que los precios barriales difieren para ingresos medios similares, podemos sospechar que las disparidades obedecen no sólo a los errores o residuos, sino a otras variables; por ejemplo, las condiciones de comodidad que imperan en cada una de las colonias, sintetizada en un *índice hedónico* (Bonin, 2010), que agrupa los servicios públicos, nivel de delincuencia, la presencia de centros comerciales, oferta de amenidades, distancia a las áreas de trabajo y el acceso a viabilidades y transporte público, por mencionar algunos de estos elementos hedónicos. A este índice le llamamos IH y actúa como una variable de control. Los atributos que lo conforman son de los barrios, no de las ciudades; por tanto, se ubica en el nivel 1. La ecuación anterior se extiende para adoptar la siguiente forma operativa:

$$P_{ij} = \beta_{0i} + \beta_{1i}X_{ij} + \beta_{2i}IH_{ij} + e_{ij}$$

Por último, la hipótesis puede aceptar otro condicionante, si las observaciones nos advierten que tanto para un ingreso como para un índice hedónico similares, los precios barriales varían de una ciudad a otra. Ello puede significar que el análisis no ha tomado en cuenta un factor que puede ser fundamental: el mercado. El vínculo entre la cantidad ofrecida (disponible) y la demandada de terrenos puede afectar el precio promedio de los barrios, en el sentido de que ahí donde son relativamente escasos tenderán a ser más altos, mientras que serán más baratos donde son relativamente más abundantes. La densidad poblacional de la ciudad es una buena *proxy* de la interacción entre oferta y demanda. Los kilómetros son un parámetro, sin no cabal, sí aproximado de los terrenos disponibles, mientras el número de habitantes puede representar la demanda por espacio. Esta variable debe referirse a la ciudad, en la medida en que los usos del suelo residencial compiten con otros (comercial, público, gubernamental, etc) por los predios. Por ello, es más razonable introducir este indicador como una varia-

ble citadina, que como una de dimensión barrial. Ello significa que la densidad poblacional es un atributo que debe ser ubicado en el nivel 2. Si llamamos D_j a la densidad poblacional de la ciudad j -ésima, el modelo multinivel queda formulado en sus dos niveles:

Nivel 1

$$P_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \beta_{2j}IH_{ij} + e_{ij}$$

En el nivel 2, la ecuación correspondiente depende de las hipótesis específicas: las discrepancias entre las ciudades obedecen a: i) los distintos promedios ciudadanos y/o ii) las diversas respuestas (pendientes) de los precios al nivel socioeconómico de las colonias. En el primer caso, el objeto de estudio subordinado son las medias; en el segundo, las pendientes. Por tal razón, podemos enunciarlas así:

Nivel 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}D_j + u_{0j}$$

y/o

$$\beta_{1j} = \gamma_{10} + \gamma_{11}D_j + u_{1j}$$

La introducción del segundo nivel, amerita algunas aclaraciones. Para facilitar la exposición, supongamos que $D_j=0$.

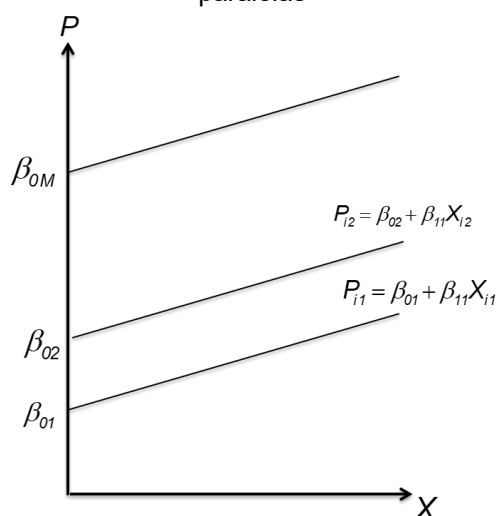
i) *Ahora los parámetros son una función:* varían de ciudad a ciudad. En esto consiste la novedad del análisis multinivel. En la recta de regresión, los coeficientes β_{0i} y β_{1i} son constantes, en virtud de que el modelo estima precisamente una recta. En el ML esta linealidad no es abandonada, en la medida en que la constancia de aquellos parámetros es válida para una ciudad en particular: la j -ésima. Pero j puede asumir los siguientes valores: 1,2, ...,m, donde este último es el total de ciudades consideradas. De esta forma, la primera metrópolis registrará los parámetros β_{01} y β_{11} ; la segunda, β_{02} y β_{12} , y la m -ésima, β_{0m} y β_{1m} . Por tanto, habrá tantas rectas de regresión como ciudades. La siguiente gráfica da cuenta de esta diversidad:

En los dos paneles tenemos distintas rectas de regresión: una para cada ciudad. La linealidad del análisis se conserva, pero a medida que pasamos

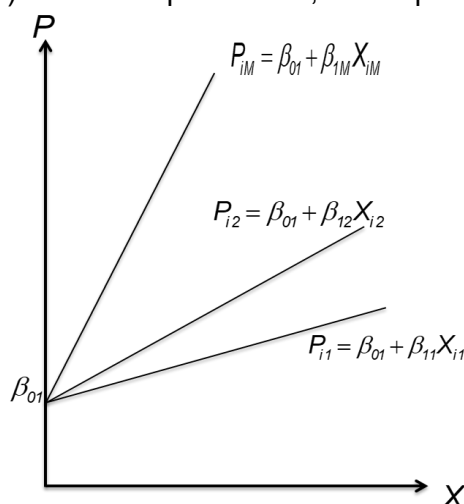
de una ciudad a otra, cambian los parámetros. En el panel (A), las rectas son paralelas y, por ello, tienen la misma pendiente. Lo que se altera es el intercepto u ordenada al origen. En (B) sucede lo contrario: todas las ciudades comparten la misma ordenada al origen, pero muestran distintas pendientes; es decir, distintas respuestas de los precios barriales al nivel socioeconómico que, como hemos dicho, es aproximado por el ingreso medio de las colonias.

Figura 1.

Rectas de regresión para distintas ciudades
(A) Diferentes interceptos, pendientes iguales y paralelas



(B) Diferentes pendientes, interceptos iguales



Fuente: elaboración propia con base en Pardo, Ruiz y San Martín, 2007.

En conclusión, las ecuaciones del nivel 2 muestran cómo es posible introducir coeficientes variables, sin abandonar la linealidad.

ii) *Efectos fijos y efectos aleatorios*. Si bien la gráfica 1 aclara nítidamente la variación de los parámetros, la forma de expresar esta variabilidad en el terreno algebraico consiste en modelarlos mediante dos componentes o efectos: un efecto fijo (γ_{00} o γ_{10}) y un efecto aleatorio (u_{0j} o u_{1j}). Es este último el que posibilita la variabilidad de los coeficientes entre ciudades de la muestra.

El efecto fijo es un valor constante en el modelo que es aplicado a cada unidad del nivel 1, con independencia de la unidad del nivel 2 en la que está anidada. Es decir, es fijo en el sentido señalado en el examen de la regresión: un valor general para todas y cada una de las colonias, ya sea de todas las ciudades o de alguna en particular. Depende de si nos concentramos en las diferencias *inter-urbes* o *intra-urbes*.

En cambio, el efecto aleatorio es una porción del coeficiente que varía entre las unidades del nivel 2; en este caso, entre las ciudades. Denota la diferencia entre el valor específico que asume el parámetro en una ciudad (una colonia) particular y un valor que es común a todas las ciudades (colonias):

$$u_{0j} = \beta_{0j} - \gamma_{00}$$

o

$$u_{1j} = \beta_{1j} - \gamma_{10}$$

Como se puede apreciar, el efecto aleatorio es el que imprime la especificidad al valor del parámetro de la unidad (ciudad) j -ésima, porque si todas comparten un valor común (las γ) y uno que sólo ella registra (las β), el resultado de la diferencia es lo estrictamente específico de esa unidad (ciudad). La inclusión del efecto aleatorio es una de las más importantes novedades del ML. En el análisis de regresión, el único componente aleatorio está representado por los errores o residuos (e), que operan en las unidades del nivel 1 (los barrios), mientras en los ML a este componente se agrega porciones aleatorias que actúan en el nivel 2.

iii) *La notación de segundo nivel* incluye como dispositivos variantes sólo a las unidades de este es-

trato (las j -ésimas ciudades) y como constantes a los efectos fijos. Estas últimas tienen un primer subíndice que indica el parámetro del nivel 1 al que corresponde, y un segundo que marca el orden secuencial en que una constante es integrada en el análisis. Por ejemplo, la constante muestra el efecto fijo del parámetro, pero al mismo tiempo el hecho de que es introducida en primer lugar (eso significa el cero). Una significa que pertenece también a aquel parámetro, pero que es introducida como segunda constante en el proceso analítico.

2. Un sentido útil para los interceptos: la media centrada.

El intercepto u ordenada al origen de cualquier recta de regresión nos dice cuánto es el valor de la variable dependiente estimada cuando su predictor es igual a cero. Nos advierte que todos los datos contienen al menos (o de más, si es negativa) esta cifra. Es, pues, un componente general de todos los registros de la muestra sobre la que se ejecuta la estimación lineal por mínimos cuadrados ordinarios (MCO). Sin embargo, para imprimirle un significado adicional, más elocuente en términos hermenéuticos, los modelos ML invocan el uso de valores escalados de la variables independientes para hacer del intercepto la media de la variable dependiente estimada:

$$XC_i = X_i - \bar{X}$$

Donde es el valor de la unidad i -ésima escalado por la gran media muestral, $\bar{X}$. Es decir, el valor escalado consiste en restar a cada una de las puntuaciones de la variable la media muestral de la misma. Con esta operación, se ubica el promedio de en punto de origen:

$$\frac{\sum X_i}{N} - \bar{X} = 0$$

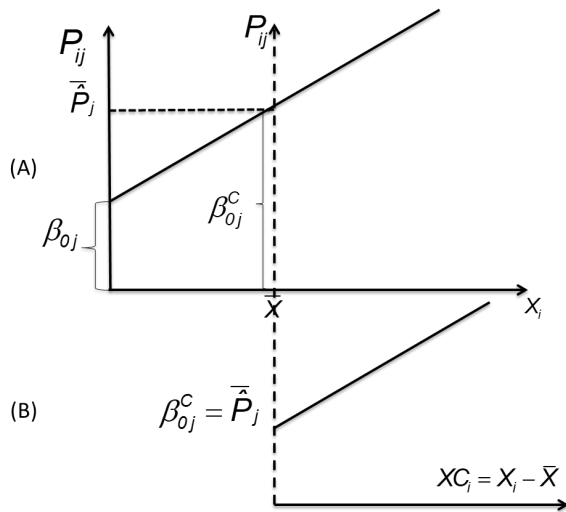
Al reestimar la regresión con los valores escalados, el resultado es un intercepto que es igual a la media condicionada de la variable dependiente. En el caso de los precios barriales estimados:

$$\beta_{0j}^e = \bar{P}_j$$

El efecto se ilustra de la siguiente forma:

Figura 2.

La media centrada de X y el intercepto



Fuente: elaboración propia

La consecuencia del reescalamiento es doble: ubica al promedio de la variable explicativa en el origen y se altera el intercepto hasta igualarse con el promedio *estimado* de la dependiente. Por tanto, en nuestro ejemplo refleja la media de los precios barriales de la ciudad *j*-ésima. Si la regresión se estima para toda la muestra, el intercepto también reflejará la gran media muestral. Otro hecho relevante es que ni los datos originales ni la pendiente de la recta de regresión se trastocan. En el panel (A) se muestra que, una vez re-escalada, la recta permanece inalterada. En este sentido, la cuantía de la respuesta de los precios al ingreso (o cualquier otra variable predictora) no se trastoca con este redimensionamiento.

Para no complicar la notación de los parámetros, usaremos la que se ha venido utilizando hasta aquí, pero bajo la advertencia de *que los interceptos serán estimados con medias centradas*. En el caso de las variables que fungen como regresores, definiremos:

$$XC_{ij} = X_{ij} - \bar{X}$$

$$IHC_{ij} = IH_{ij} - \overline{IH}$$

$$DC_j = D_j - \bar{D}_j$$

Es decir, los ejercicios que se desarrollarán en los siguientes modelos se refieren invariablemente a datos re-escalados por su media.

II. Modelo 1. Análisis de varianza (ANOVA) y la diferencias entre medias.

¿Existen diferencias entre los precios barriales que sean estadísticamente significativas?. Esta sección está dirigida a contestar esta pregunta. Y por eso es muy cercana al análisis de varianza (ANOVA) de una dirección. La estrategia metodológica para su respuesta se concentra en el modelo nulo o vacío, que consiste en desplegar el análisis multinivel *sin variables independientes o predictoras*.

1. Las ecuaciones multinivel del modelo nulo

Al no contener variables predictoras o explicativas, el modelo nulo se expresa con las siguientes ecuaciones multinivel:

$$P_{ij} = \beta_{0j} + e_{ij} \quad \text{Nivel 1}$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad \text{Nivel 2}$$

Como antes, P_{ij} es el precio barrial de la *i*-ésima colonia, ubicada en la *j*-ésima ciudad de la muestra. β_{0j} es el intercepto que refleja la media de los precios barriales en ese centro urbano. Por su parte, e_{ij} es el error o residuo y muestra la parte no explicada de aquel precio, que indica la desviación entre esa cotización particular y el promedio ciudadano. Da cuenta de las diferencias intergrupales, ya sea porque alude distintas colonias de una sociedad y/o porque distingue entre las que se encuentran en urbes diferentes. La incorporación del subíndice “*i*” significa que cambia con cada colonia, mientras la del correspondiente a “*j*” que muta al inspeccionar otra ciudad.

El foco de atención, sin embargo, debe reposar en el nivel 2. En primer lugar, el intercepto β_{0j} es una función que representa una media urbana que varía de una ciudad a otra. Por ejemplo, si $j=1$ representa a Guadalajara, el intercepto será el precio barrial promedio de esta entidad jalisciense; si $j=2$ corresponde a Monterrey, será la media regiomontana. Se trata, por tanto, de la media de una ciudad determinada.

En segundo lugar, el intercepto contiene un efecto fijo (γ_{00}) y un efecto aleatorio (u_{0j}). El primero es fijo en el sentido de que no cambia para ningún barrio, aunque cambiemos de ciudad: *todas de las colonias de todas las ciudades incluirán este componente*. Por tal razón, no incorpora el subíndice “j”; es un artificio para indicar que el dato no se modifica cuando se pasa de una urbe a otra. *Es importante recordar que representa la gran media o media muestral; es decir, el promedio del precio barrial de todas las colonias en la muestra*. En cambio, el componente aleatorio incorpora “j”, pero excluye el subíndice “i”. Ello implica que no varía entre las colonias de una ciudad y, por tanto, que es general a todos sus barrios, pero se altera cuando es otra la urbe considerada. Es este efecto el que permite que cambie el intercepto (y por tanto, la media ciudadina) de un conglomerado urbano a otro, en la medida de que la otra parte que lo integra (γ_{00}) permanece inalterada cuando analíticamente realizamos este movimiento. Lo importante es que da cuenta de la diferencia entre la media grupal (el promedio de los precios barriales en la ciudad j-ésima) y la gran media (el promedio muestral) de todas las colonias en todas la ciudades:

$$u_{0j} = \beta_{0j} - \gamma_{00}$$

Las variaciones y, por tanto, la varianza de este efecto aleatorio dan cuenta de la diferencias entre los medias de los precios barriales entre ciudades.

2. El modelo mixto en la modalidad nula.

El modelo mixto es una combinación de los dos niveles. Su ecuación es el objetivo de los cálculos que un investigador debe de realizar. El procedimiento consiste en sustituir la o las ecuaciones del nivel 2 en la del nivel 1:

$$P_{ij} = \gamma_{00} + (u_{0j} + e_{ij})$$

De esta forma, el precio barrial de una colonia i-ésima está determinado por un efecto fijo y dos efectos aleatorios, que son los que se encuentran entre paréntesis. El primero de estos últimos responde a la variación entre ciudades (intergrupar),

mientras el segundo a lo que ocurre dentro de las ciudades (intragrupal).

3. Una estimación hipotética con fines estrictamente expositivos del modelo nulo.

La estimación de los modelos ML puede realizarse mediante distintos paquetes estadísticos¹. En esta sección nos restringimos a exponer la forma en que se interpretan los resultados. Los datos son hipotéticos, y no, el resultado de una investigación original. La intención es pedagógica y nada más. Por tal razón, *no* deben asumirse verdaderos.

Generalmente, los software entregan tres tipos de estimaciones, que son las relevantes para la interpretación de los distintos modelos que seguiremos en estas líneas: la de los parámetros de los efectos fijos (cuadro 1); la de los parámetros de covarianza (cuadro 2), que reporta la varianza de los efectos aleatorios, y el estadístico de prueba -2LL, cuya utilidad reside en sustituir la información del correspondiente a la Wald Z, cuando ésta arroja dudas a causa de que la muestra es pequeña.

Esta utilidad, sin embargo, cobra sentido cuando se comparan dos modelos. Por tanto, haremos caso omiso de la misma en el primer modelo (el nulo), para volver a él cuando se introduzca el segundo ejercicio.

En el rubro “Modelo nulo” del cuadro 1 se muestran sus efectos fijos y los estadísticos de prueba de la regresión correspondiente. El intercepto, como se ha dicho, representa la gran media o media muestral. El valor estimado es: $\gamma_{00}=918.3$. Este valor refleja el promedio general (para todas las colonias de la muestra) del precio barrial, con independencia de la ciudad en la que se encuentren. El valor de la prueba t y el nivel de significancia asociado (sig.<0.05) nos dicen que este parámetro es estadísticamente significativo; por tanto, se rechaza la hipótesis nula: $H_0: \gamma_{00}=0$. En términos más prácticos, podemos asumir que el valor estimado es válido.

¹ Para el uso de SPSS puede consultarse Pardo, Ruiz y San Martín (2007); para el uso de Stata, Luis Gandós (2014).

Cuadro 1.
Ejemplo de la estimación de los parámetros de efectos fijos

Modelo	Parámetro		Desviación Estándar	t	Sig
	Nombre	Valor estimado			
Nulo	Intercepto : γ_{00}	918.3	96.6	9.50621118	0.000
Modelo 2	Intercepto : γ_{00}	891.34	89.7	9.93690078	0.000
	X: Ingreso barrial : γ_{10}	0.411	0.0925	4.443243243	0.000
Modelo 3	Intercepto : γ_{00}	906.47	72.5176	12.5	0.000
	X: Ingreso barrial : γ_{10}	0.2783	0.0696	3.999	0.000
	IH: : γ_{20}	0.23	0.0434	5.300	0.000
Modelo 4	Intercepto : γ_{00}	921.880	77.5340614	11.89	0.000
	X: Ingreso barrial : γ_{10}	0.284	0.088243571	3.22	0.002
	IH: : γ_{20}	0.248	0.03488608	7.119	0.000
Modelo 5	Intercepto : γ_{00}	1149.216	230.766	4.98	0.000
	X: Ingreso barrial : γ_{10}	1.060	0.301	3.52	0.000
	IH: : γ_{20}	0.249	0.034	7.32	0.000
	D: : γ_{01}	0.192	0.069	2.78	0
	X*D: : γ_{11}	0.035	0.012	2.88	0.018

Fuente: elaboración propia con datos ficticios y con fines expositivos

Cuadro 2.
Ejemplo de la estimación de los parámetros de covarianza (efectos aleatorios)

Modelo	Parámetro		Desviación es- tándar	Wald Z	Sig.
	Nombre	Estimación			
Nulo (-2LL= 4602.4)	Varianza de residuos	1808.49	185.79	9.734	0.000
	Varianza de U_{0j}	913.28	430.02	2.124	0.035
Modelo 2 (-2LL=4324.3)	Varianza de residuos	1628.65	167.63	9.715743005	0.000
	Varianza de U_{0j}	230.07	147.48	1.560008137	0.066
Modelo 3 (-2LL=4128.7)	Varianza de residuos	1485.43	112.0740908	13.254	0.000
	Varianza de U_{0j}	248.71	43.23887344	5.752	0.000
Modelo 4 (-2LL=3898.9)	Varianza de residuos	1290.690	89.013	14.5	0.000
	Varianza de $U_{0j} : UN(1,1)$	259.902	56.834	4.573	0.000
	Covarianza ($U_{0j}U_{1j}$): UN(2,1)	5.796	3.067	1.89	0.062
	Varianza de $U_{1j} : UN(2,2)$	1.683	0.652325581	2.58	0.000
Modelo 5 (-2LL=3714.1)	Varianza de residuos	1276.493	100.7492135	12.67	0.000
	Varianza de $U_{0j} : UN(1,1)$	213.120	36.18329355	5.89	0.000
	Covarianza ($U_{0j}U_{1j}$): UN(2,1)	5.269	0.564734624	9.33	0.000
	Varianza de $U_{1j} : UN(2,2)$	1.467	0.230283014	6.37	0.000

Fuente: elaboración propia con datos ficticios y con fines expositivos

La varianza de los residuos, σ_e^2 , indica cuánto varían los precios barriales dentro de cada ciudad. Su valor ($\sigma_e^2 = 1808.49$) es estadísticamente significativo, como nos lo da a saber el estadístico Wald Z y el nivel de significancia vinculado a este criterio de prueba (sig. 0.05). Esta varianza es otra cosa que la suma de cuadrados no explicada, entre sus grados de libertad. El coeficiente Wald Z es una especie de prueba t, que normalmente se usa como criterio para contrastar la hipótesis nula de que un parámetro particular sea igual a cero. Sin embargo, en lugar de dividir el parámetro estimado entre su desviación estándar, la operación se realiza entre la varianza estimada y su error típico correspondiente. La confianza que podemos depositar en él es limitada, en la medida en que sólo es contundente cuando nos encontramos con un tamaño de muestra lo suficientemente grande. Como veremos, es más conveniente examinar el cambio entre los valores de -2LL de dos modelos alternativos.

El segundo efecto aleatorio, u_{0i} , registra una varianza hipotética de: $\sigma_{u0}^2 = 913.28$; también es significativa, según nos informa la prueba Wald Z y el nivel de vinculado a ella (sig. <0.05). Por tanto, esos datos avalan que, efectivamente, existen variaciones entre los precios barriales que obedecen a las diferencias entre las ciudades y no solamente a discrepancias dentro de la mismas, como atestigua la varianza de los residuos. En este sentido, queda contestada la pregunta que inauguró este apartado: las colonias de la muestra exhiben diferenciales *significativos* en precios barriales; una parte de ellos es explicada por las discrepancias entre las cotizaciones de cada colonia y la media ciudadina, mientras que otra porción responde a las que se registran entre esta última medida central y la gran media (muestral).

Sin embargo, abre una nueva cuestión: ¿cuánto de esa variación total puede ser atribuida a las diferencias interciudades y cuánto a las intraciudades?. La respuesta nos la otorga el *coeficiente de correlación interclase (CCI)*, que mide la proporción de la varianza total explicada por el factor (la ciudad) (Pardo et. al, 2007).

$$CCI = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2} = \frac{913.8}{(913.8 + 1808.49)} = 0.3356$$

El resultado nos advierte que 33.56% de las variaciones en los precios barriales es causadas por las

discrepancias entre ciudades, mientras que el resto (66.44%) por las que operan dentro de ellas. Recordemos que las primeras dan cuenta de las diferencias entre las medias grupales (en este caso, las de las ciudades) y se miden mediante sus distancias con relación al promedio general o gran media.

III. El modelo 2: los determinantes de las diferencias intra e interciudadinas

El modelo nulo avaló que existen diferencias estadísticamente significativas en los precios barriales, que además pueden ser atribuidas a variaciones dentro y entre las urbes. La investigación cuenta, entonces, con un objeto de estudio respaldado por técnicas cuantitativas y, por tanto, con una pregunta que puede guiar la auscultación científica sobre un terreno firme, al inquirir adecuadamente sobre las causas que explican los diferenciales de precios intra e intraciudadinos.

La hipótesis central adelantaba que una posible causa son las disparidades socioeconómicas que permean a las colonias, aproximadas por lo que hemos llamado ingreso barrial. Es la hora de introducir esta variable independiente en el análisis y, por esta vía, someter a prueba aquella hipótesis rectora.

Sin embargo, la forma de hacerlo tiene que ser liderada por las nuevas indagaciones que suscitan los resultados del modelo nulo: si la variación de los precios barriales se explica tanto por la diferencias entre ciudades (variación explicada) como dentro de la ciudad (variaciones no explicadas), ¿qué es lo que determina esas diferencias? Según la hipótesis principal, el nivel socioeconómico (ingreso barrial) es un candidato protagonista para contestar esa inquisición. Imaginemos por un momento que la hipótesis es correcta, ¿cómo se manifiesta este acierto en el campo estadístico? En primer lugar, en la reducción de la varianza no explicada: una porción de esta pasa a ser “explicada” por la nueva variable predictora. Pero, además, en que una distinción de una unidad de ingreso barrial (uno, cien, mil pesos, etc) provoca una variación en esas cotizaciones equivalente a la cuantía de la pendiente de la regresión. Esto significa que, si se logra fijar esa cuantía y se generaliza para todas las ciudades, las diferencias entre las medias ciudadinas de los precios estarán causadas por las discrepancias en los ingresos barriales promedio de las urbes.

1. Las ecuaciones multinivel del modelo 2.

Ya se adelantó que el estatus socioeconómico, aproximado por el ingreso barrial, es una variable del nivel 1, porque es considerado como un atributo de las colonias, y no de las ciudades que las aglutinan. Por tal razón, la principal modificación opera en la ecuación que da cuenta de este nivel:

$$P_{ij} = \beta_{0j} + \beta_{1j}XC_{ij} + e_{ij} \quad \text{Nivel 1}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}; \beta_{1j} = \gamma_{10} \quad \text{Nivel 2}$$

Al reescalar X hasta ubicar su media en el origen, el parámetro β_{0j} equivale a la media de los precios barriales en una ciudad j-ésima. Pero se trata ahora de una media condicionada a que x_{ij} asuma un valor igual a su promedio $\bar{x}$. Si no se hubiera centrado la media, la condición habría sido que $x_{ij}=0$; es decir, que el promedio de ingreso barrial fuera nulo. Lo que resulta no sólo irrealista, sino absurdo, pero estadísticamente correcto, como apunta Hayes (2002). Por tanto, cobra sentido trabajar con medias centradas. Lo mismo sucede con γ_{00} : sigue representando la gran media, pero ahora sujeta a la misma condición.

Al introducir la nueva variable explicativa, también aparece un nuevo parámetro: β_{1j} , que representa la pendiente de la recta de regresión y, por ello, la cuantía de la respuesta de los precios barriales al estatus socioeconómicos; o mejor, el indicador que nos dice en cuánto se distinguen los precios de dos colonias por cada peso (cientos; miles, etc.) de diferencia en sus respectivos ingresos barriales. Sin embargo, este parámetro es introducido también en el nivel 2, con dos características especiales para este modelo: a) sin efectos aleatorios y con un solo efecto fijo: γ_{10} ; y b), sin que este componente contenga a "j" como uno de sus subíndices (ni a "i", por supuesto, por tratarse del nivel 2). Con ello se asegura no sólo la fijeza de la pendiente en el nivel 1, sino también que su valor sea general para todas las ciudades. Tal acto garantiza que el monto de la respuesta de los precios a los ingresos barriales sea común, constante, para todas las colonias; y con ello las variaciones en las cotizaciones de los terrenos respondan únicamente a las diferencias entre los ingresos barriales. Vale decir: la respuesta de los precios corresponde a las variaciones de XC_{ij} , y no a las de β_{1j} cuando nos movemos de ciudad a otra, *aún cuando intuitivamente podamos suponer que este parámetro cambia al realizar esta mudan-*

za analítica. La intención es aislar la asociación entre *niveles* de ingreso y *niveles* de precio de las posibles variaciones en las pendientes ciudadinas. Con ello se consigue explicar las diferencias interurbanas de las medias de los precios barriales mediante discrepancias en las del ingreso barrial de las ciudades: si uno de estos conglomerados registra un promedio de ingreso mayor que otro, tal diferencia deberá expresarse en una cotización media de los terrenos distinta.

2. El modelo mixto de la relación entre precios barriales y estatus socioeconómico.

El modelo mixto, como es costumbre, se articula al sustituir las ecuaciones del nivel 2 en la del nivel 1:

$$P_{ij} = \gamma_{00} + \gamma_{10}XC_{ij} + (u_{0j} + e_{ij})$$

Los efectos aleatorios son los mismos que en el modelo nulo y se expresan entre paréntesis. u_{0j} sigue siendo la diferencia entre la media ciudadina (β_{0j}) y la gran media (γ_{00}), pero ahora condicionadas a que $XC_{ij}=0$. La novedad reside en la pendiente de esta ecuación, suministrada por γ_{10} , que es asumida como respuesta de los precios a los ingresos barriales que es aplicada a todas las colonias de la muestra.

3. Estimación (hipotética) e interpretación de resultados del modelo 2.

El cuadro 1 muestra unos resultados hipotéticos de la estimación del modelo 2 que, igual que el anterior, tiene propósitos exclusivamente expositivos. Como se puede apreciar, $\gamma_{00}=891.34$, lo que refleja que la gran media de los precios barriales es menor a la que registraba el modelo 1, a pesar de que tanto la prueba t como el nivel de significancia (sig.<0.05) nos confirman que el parámetro es estadísticamente significativo. La razón de esta disminución reside en que *ahora es una gran media condicionada, mientras en el modelo nulo se trataba de otra de índole incondicionada*. Más específicamente, representa la media de todas las colonias, con independencia de la ciudad en la que se encuentre, cuando el ingreso barrial de cada una de ellas es igual al promedio de esta variable explicativa. ¿Cuál es el propósito de esta restricción? Anular los efectos que sobre la media muestral incondicionada tiene el ingreso barrial.

El artificio suprime las diferencias en este indicador al usar como contrafactual un nivel de ingreso y una respuesta de los precios barriales a esta variable (pendiente) similares en todas las colonias. Recordemos que $\beta_{1j} = \gamma_{10}$ significa que aquella respuesta es igual para todas las colonias; si hacemos que $XC_{ij} = 0$, entonces . Es decir, el efecto del ingreso barrial queda anulado, porque se está condicionando a que todas las colonias registren el mismo ingreso barrial no centrado, que equivale precisamente a la media muestral de esta variable.

La atención debe recaer en el nuevo parámetro introducido: $\gamma_{10} = 0.411$. Es la pendiente media (general para toda la muestra de colonias), que también es significativa. El signo positivo nos advierte que a mayor ingreso barrial (estatus socioeconómico), mayor serán el precio barrial. Pero también que, con independencia de la ciudad en la que se ubiquen, una diferencia en una unidad de ingreso barrial se traducirá en un diferencial del precio de los terrenos igual a 0.411, medido en la misma unidad. Por ejemplo, si en una colonia el ingreso barrial es de 600 dólares y en otra es de 800, el precio barrial en esta última será 82.2 mayor que en la primera.

¿Esto significa que no podemos conocer los determinantes de las diferencias entre las medias citadinas de los precios barriales? Todo lo contrario: el modelo 2 está diseñado sobre todo para ese propósito. Fijar la pendiente persigue uniformar la respuesta de los precios a la diferencia en una unidad de ingreso barrial para todas las colonias y todas las

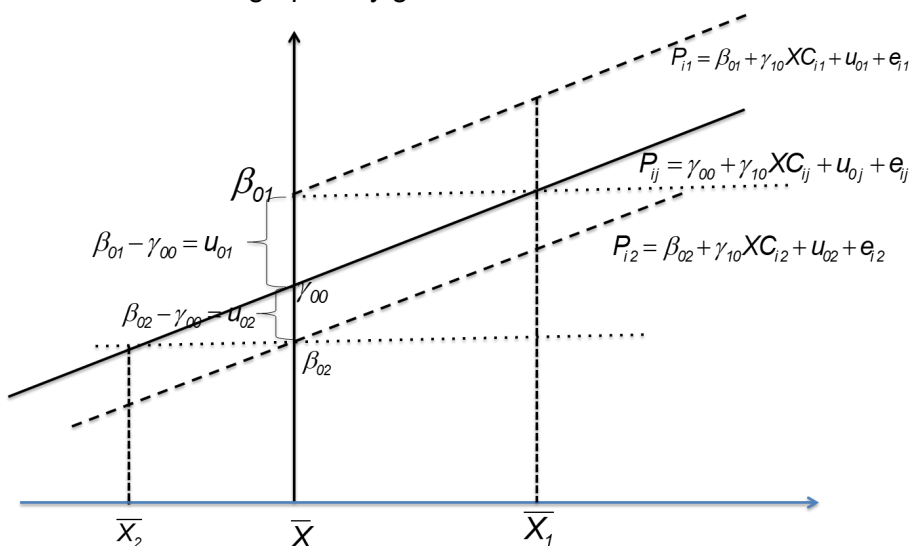
ciudades, con finalidad de que estas cotizaciones se distingan únicamente por la diferencia de ingresos barriales. Es decir, fijamos β_{1j} para que varíe solamente XC_{ij} . Y éste cambia, tanto cuando lo hace “i” (la colonia) como cuando muta “j” (la ciudad).

Sin embargo, sabemos que el intercepto β_{0j} refleja la media de los precios barriales de una ciudad j-ésima cuando $XC_{ij}=0$. Ello implica que $X_{ij}=X'$; es decir, que el ingreso barrial no centrado asume un valor equivalente a la gran media muestral. Por tanto, se tiene una media *estimada* de los precios barriales para cada una de las ciudades, a medida que j asume valores desde 1 hasta la m-ésima ciudad de la muestra y que, al hacerlo, varía la media del ingreso barrial citadina.

$$\hat{\bar{P}}_j = (\beta_{0j} | X_i - \bar{X} = 0)$$

Lo que esta ecuación expresa es que el intercepto de la recta de regresión de cada ciudad j-ésima refleja el promedio de los precios barriales si el ingreso está centrado por su gran media muestral. Es decir, el escalamiento no sólo conduce a que el intercepto de la regresión total (γ_{00}) sea igual la media general de las cotizaciones barriales, sino también que los particulares de cada ciudad equivalgan a sus respectivas medias. Esto se puede ilustrar mediante la gráfica siguiente:

Figura 3.
Medias grupales y gran media centrada



Fuente: elaboración propia

En ella, la línea gruesa y continua es la recta de regresión muestral. Al centrar con el ingreso barrial promedio (X'), el intercepto γ_{00} refleja la gran media. Si se estima una regresión para la ciudad 1, el intercepto β_{01} corresponderá a la media del precio barrial de esa ciudad; lo mismo sucede con β_{02} , si el ejercicio se despliega para una segunda urbe. Las dos rectas de regresión son paralelas porque registran la misma pendiente que la general: γ_{10} . Este es el rasgo característico del modelo 2. Sin embargo, se aprecia que los interceptos ciudadanos acusan distancias distintas con relación a la gran media muestral.

También se aprecia en la gráfica que las medias ciudadinas de los precios varían con los promedios de ingreso barriales de cada urbe. La primera ciudad registra uno que equivale a X'_1 ; al ubicar el valor que le corresponde en la recta de regresión muestral (la general), el resultado es el intercepto β_{01} , que es precisamente la media de los precios barriales para este núcleo urbano. En cambio, si pasamos a la segunda, el promedio de ingreso ciudadano varía y es equivalente a X'_2 ; la consecuencia es una mutación en el intercepto: β_{02} . Por tanto, el modelo 2 muestra las diferencias entre las medias interurbanas de los precios barriales como una función de las discrepancias entre los promedios de los ingresos barriales de las ciudades.

El cuadro 2 exhibe los parámetros de covarianza para los efectos aleatorios. El primer foco de atención debe de reposar en la varianza de los residuos: $\sigma^2_{\epsilon 1}=1628.65$. Este valor es sustancialmente menor al que acusó el modelo nulo: $\sigma^2_e=1808.49$. El significado es que la introducción del ingreso barrial como variable independiente disminuyó esta varianza no explicada porque una porción de la original pasó a ser explicada por ella. La cantidad absoluta de este cambio es: $\sigma^2_e - \sigma^2_{\epsilon 1} = 1808.49 - 1628.65 = 179.84$. Pero esta cifra, por sí misma, no dice nada; la relativa es más elocuente:

$$\frac{\sigma^2_e - \sigma^2_{\epsilon 1}}{\sigma^2_e} = 1 - \frac{\sigma^2_{\epsilon 1}}{\sigma^2_e} = 1 - \frac{1628.65}{1808.89} = 0.099$$

Es decir, los distintos ingresos explican, en promedio, 9.9% de las diferencias intra ciudadinas de los precios de los precios barriales. Por tanto, es necesario incorporar otras variables del nivel 1 para reducir todavía más el 90.1% restante de esta varianza no explicada. De ello nos ocuparemos en el siguiente modelo.

El segundo efecto aleatorio corresponde al factor ciudad: u_{0j} ; marca la diferencia entre la cotización media ciudadina y la gran media muestral de los precios barriales. La gráfica 3 muestra cómo este factor varía con los interceptos ciudadanos, y éstos con sus respectivos ingresos medios. Por tanto, su varianza está ahora condicionada a que $XC_i=0$. Su valor refleja una reducción con relación al modelo nulo: de 913.28 a 230.07, que equivale a una baja de 74.8%. La interpretación es que la diferencia entre los ingresos medios de las ciudades es responsable de este porcentaje de variaciones. Ciudades más ricas tenderán a ostentar, en promedio, mayores precios barriales.

Sin embargo, el lector notará que la varianza del efecto ciudad no es significativa (sig.>.05). Por tanto, su verdadero valor no es distinto de cero. Esta es la ocasión de introducir una prueba alternativa, más robusta, que es aconsejada por todos los analistas, en virtud de que la Wald Z no es del todo robusta, especialmente cuando se trata de muestras pequeñas. La clave consiste en comparar el valor que asume el estadístico -2LL en dos modelos. La diferencia entre ambos indicadores se distribuye como una χ^2 con grados de libertad equivalentes a la diferencia de parámetros estimados en uno y otro modelo. Según el cuadro 1, en el modelo nulo fue de -2LL=4,602.4, mientras que en el modelo 2 este indicador alcanzó la cifra de 4,324.3. La diferencia entre ambos es igual 278.1. Por otro lado, la diferencia de parámetros estimados es uno (γ_{10}). Por tanto, el valor de p para $\chi^2(1)=278.1$ es $p<0.001<0.05$. Por ello, la varianza estimada es estadísticamente significativa y distinta de cero. Se recomienda realizar este tipo de ejercicio para todos los modelos, a pesar de que por razones de espacio no se harán con los siguientes, los cuales se sujetarán exclusivamente a la prueba Wald Z.

IV. El Modelo 3: la incorporación de una variable de control: el Índice Hedonístico.

El modelo 2 constató que el nivel socioeconómico de las colonias, aproximado por su ingreso barrial, es una variable decisiva en los precios barriales. Sin embargo, éstos pueden estar influidos por las condiciones de habitabilidad, especialmente aquellas que definen la comodidad y el placer vivir en ellas. Ahí pueden imperar distinciones importantes en los servicios públicos, el grado de delincuencia, la presencia de centros comerciales, la oferta de amenidades, la distancia a las áreas de trabajo y el acceso a viabilidades y transporte público. La lista puede ser extendida o restringida según los crite-

rios de cada investigador; lo importante es que todas estos atributos pueden ser transformados en elementos medibles de un índice de placer, que recorra una escala determinada en la que los niveles más bajos señalen carencias o insatisfacción, mientras los más altos una vida más placentera. A este indicador le llamamos Índice Hedonístico que, como se dijo en la primera sección, identificamos con la siglas IH. Una vez que se ha centrado por su media, estas siglas se convierten en IHC.

El propósito de incluir esta variable es controlar la respuesta de los precios a los ingresos barriales: ¿se mantiene esta relación cuando se elimina la influencia de las condiciones hedónicas? ¿hasta qué punto inciden estos atributos en los precios barriales? Es decir, se debe de asegurar que las variaciones de precios reflejen los distinguos entre niveles de ingreso de la forma más “pura” posible: libre de otras injerencias que conduzcan a errores de apreciación y estimación.

1. Las ecuaciones multinivel del modelo 3.

El Índice Hedónico Centrado (IHC) es un atributo de las colonias, y no de las ciudades. Por tanto, debe ubicarse en el nivel 1; sin embargo, el propósito central del modelo impone nuevas condiciones en las ecuaciones del nivel 2:

$$P_{ij} = \beta_{0j} + \beta_{1j}XC_{ij} + \beta_{2j}IHC_{ij} + e_{ij} \quad \text{Nivel 1}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}; \beta_{1j} = \gamma_{10}; \beta_{2j} = \gamma_{20} \quad \text{Nivel 2}$$

Como se puede apreciar, la variable IHC es acompañada del parámetro β_{2j} , que es la pendiente de los precios a este índice: nos informa en cuanto difieren las cotizaciones barriales por cada punto de diferencia en este indicador. Sin embargo, en el nivel 2 se impone que $\beta_{2j} = \gamma_{20}$. Es decir, se fija esta relación en el mismo sentido que con la pendiente de los ingresos: la relación entre precios y la variable predictora, suministrada por la pendiente, se asume que es general, común, a todas las ciudades de la muestra. La idea es la misma que en el modelo 2: contrastar las variaciones de los precios con movimientos en el índice, sin la influencia de las posibles discrepancias en la sensibilidad de aquella variable a este nuevo predictor; es decir, de las variaciones en la pendiente.

Como antes, el modelo mixto se configura al sustituir las expresiones del nivel 2 en las ecuación del nivel 1:

$$P_{ij} = \gamma_{00} + \gamma_{10}XC_{ij} + \gamma_{20}IHC_{ij} + (u_{0j} + e_{ij})$$

El parámetro γ_{00} continúa representando la gran media, pero ahora sujeta a un doble condicionamiento: $XC_{ij}=0$ y $IHC_{ij}=0$. Es decir, refleja el promedio de los precios barriales a lo largo de la muestra, cuando la colonias y ciudades registran el promedio muestral del ingreso (X') y el del Índice Hedonístico: IHC' . La intención es homogenizar las colonias y ciudades en estos rubros; al combinarse con pendientes que son iguales para todas las urbes ($\beta_{1j} = \gamma_{10}$; $\beta_{2j} = \gamma_{20}$) se anulan los efectos de estos dos predictores y queda una gran media exenta de esas influencias.

Por su parte, γ_{10} ahora está condicionada a que $IHC_{ij}=0$; es decir, a que todas colonias registren el mismo Índice Hedónico, igual a su promedio muestral: IHC' . Nos informa en cuánto se distinguen los precios por cada unidad de diferencia en los ingresos barriales, *con independencia de las características hedónicas de cada colonia*. Lo contrario sucede con γ_{20} : nos informa en cuanto varían aquellas cotizaciones por cada punto de diferencia en el IHC , cuando $XC_{ij}=0$: independientemente de los ingresos barriales. Libre de esta influencia.

2. Estimación hipotética del modelo 3

El cuadro 1 muestra los datos hipotéticos de los efectos fijos: γ_{00} γ_{10} γ_{20} . Los tres son estadísticamente significativos. La gran media no ha sufrido una alteración significativa, pero su ligero ascenso con relación al valor registrado en el modelo 2 es una indicador de que un IH relativamente bajo estuvo empujando hacia abajo el precio barrial de la muestra. Una vez que se libera de esta influencia, la gran media aumenta. La pendiente de los ingresos, γ_{10} , es positiva, lo que nos informa que por cada unidad de diferencia en los ingresos barriales, los precios barriales se distinguen entre sí en 0.2783 dólares. Esta respuesta sufrió una reducción al incorporar la variable de control, lo cual es signo de que la correspondiente al modelo 2 estaba impregnada todavía de las características que definen el IH. Al despojarla de esta injerencia, exhibe una relación más cercana entre las dos variables de la hipótesis central. Por su parte, el valor de γ_{20} y, sobre todo, su nivel de significancia advierte que los atributos sintetizados en el IH provocan un diferencial en la cotizaciones de los terrenos coloniales de 0.23 por ca-

da punto de variación que experimenta la covariable de control, cuando soslayamos las discrepancias entre los ingresos barriales. El hecho de que su valor sea inferior a γ_{10} implica que el nivel socioeconómico es todavía el actor protagónico a la hora de explicar las disimilitudes entre los precios barriales.

El cuadro 2 muestra los datos hipotéticos de los efectos aleatorios. La varianza de los residuos de este modelo 3, (σ^2_{e2}), refleja la varianza no explicada que resta, una vez que se ha descontado la porción explicada por la acción conjunta de los ingresos barriales y el IH. Por tanto, si queremos saber cuánto es esta ración, podemos aplicar el mismo procedimiento:

$$1 - \frac{\sigma^2_{e2}}{\sigma^2_\epsilon} = 1 - \frac{1485.43}{1808.49} = 0.1786$$

Por tanto, X y IH dan cuenta conjuntamente de 17.86% de la varianza residual. Si atendemos a que en el modelo 2, los ingresos barriales explicaban 9.9% de esta medida de dispersión, es posible que concluir que el IH es responsable de 7.96% (17.86-9.9) de las diferencias intracitadinas de los precios barriales.

Por último, la varianza de u_{0j} , que muestra las discrepancias interurbanas de los precios barriales, ha experimentado un aumento de 8.1%, con relación al modelo 2, lo que denuncia disimilitudes en la infraestructura urbana, tanto pública como privada, que se traducen en disparidades entre las ciudades, en términos de habitabilidad y comodidad, aún si el ingreso promedio fuera igual para todas y cada una de ellas. Es decir, el carácter más o menos “amigable” de las zonas habitacionales y, por tanto, de las urbes, se plasma en mayores o menores precios barriales.

V. El modelo 4: las pendientes aleatorias y la disparidad de las respuestas urbanas.

Los modelos anteriores también se les conoce como de interceptos aleatorios, porque estos parámetros son los únicos que varían entre colonias y entre ciudades. Su examen demandó fijar las pendientes o respuesta de los precios barriales a las dos covariables introducidas. Sin embargo, el análisis multinivel no es ajeno a la idea de que, en los hechos, el vínculo entre la variable independiente y sus predictoras cambia de una ciudad a otra, de tal forma que asignar a todas ellas la pendiente media como vínculo común y generalizado, levanta sospechas sobre su apego a la realidad. El análisis de pendien-

tes variables, que en la literatura agrupa a los llamados *modelos de coeficientes aleatorios*, se realiza en dos etapas que recuerdan mucho la estructura analítica de las medias, seguida hasta aquí: en la primera, se verifica que, efectivamente, los datos muestran variabilidad de aquellos movimientos; de resultar cierta esta intuición, se prosigue con las segunda fase, abocada a averiguar cuáles son las causas o determinantes de esa variabilidad. La primera fue la tarea del modelo nulo; la segunda, la de los subsecuentes. La diferencia entre las dos se reproduce en este nuevo ejercicio: en la instancia preliminar se conserva el modelo con el mismo tipo de variables (en el nulo no existían variables predictoras), para después pasar a la introducción de predictoras que explican las diferencias entre las respuestas de los precios a los ingresos barriales que acusan las distintas ciudades.

El modelo 4 que se despliega en este apartado se concentra en la primera de estas fases. El objetivo consiste en atestiguar la variación de las pendientes ciudadinas, con el propósito de verificar que cada una de las urbes responde de forma distinta cuando queremos determinar la reacción de los precios a los ingresos barriales. La forma de hacerlo consiste en convertir la pendiente en una variable que contiene un efecto fijo y otro aleatorio. Si la varianza de este último es estadísticamente distinta de cero, entonces los datos confirman la variación de las pendientes entre las ciudades; si no es así, tal variación no existe y, por tanto, el análisis se detiene: las diferencias en los precios barriales obedece sólo a las diferencias entre las medias urbanas.

1. Las ecuaciones del modelo multinivel de coeficientes aleatorios: primera fase.

En la medida en que este modelo parte del modelo 3, sin introducción de nuevas variables, porque el objetivo no es explicar las causas de la variabilidad de las pendientes, sino avalar o rechazar que tales discrepancias son ofrecidas por los datos, el nivel 1 no sufre ningún tipo de modificación. Es en el nivel 2, donde se experimentan las principales alteraciones:

$$P_{ij} = \beta_{0j} + \beta_{1j}XC_{ij} + \beta_{2j}IHC_{ij} + e_{ij} \quad \text{Nivel 1}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}; \beta_{1j} = \gamma_{10} + u_{1j} \quad \beta_{2j} = \gamma_{20}$$

Nivel 2

Como se puede apreciar, el único cambio lo registra β_{1j} , la pendiente de los precios a los ingresos barriales, porque este parámetro es el foco de atención: es ahí donde se debe comprobar si varía cuando pasamos de la ciudad $j=1$ a la $j=2 \dots j=m$. Por tanto, debe modelarse como una función y abandonar el estatus fijo que le imponían los modelos anteriores. La ecuación $\beta_{1j} = \gamma_{10} + u_{1j}$ cumple con esta responsabilidad: por un lado, contiene un efecto fijo (γ_{10}) y otro aleatorio (u_{1j}). El primero, como todo componente fijo, es un promedio: representa la pendiente media de toda la muestra (la gran pendiente) y es común a todas las ciudades. Por tanto, la única fuente de variación debe ser el componente aleatorio. En él reposa la eventual variabilidad de las pendientes:

$$\beta_{1j} - \gamma_{10} = u_{1j}$$

Si éstas cambian cuando se transita de la urbe j -ésima a la m -ésima, el parámetro β_{1j} muta, y estas mutaciones se reflejan únicamente en variaciones de u_{1j} . La expresión nos dice también que representa la diferencia entre la pendiente de la ciudad j -ésima y la gran pendiente. Por tanto, mide la variabilidad, en el sentido expuesto en el recuadro 2, que consigna el ANOVA.

2. El modelo mixto de coeficientes aleatorios: primera fase.

El modelo mixto se obtiene al sustituir las ecuaciones del nivel 2 en la que da cuenta del nivel 1:

$$P_{ij} = \gamma_{00} + (\gamma_{10} + u_{1j})XC_{ij} + \gamma_{20}IHC_{ij} + (u_{0j} + e_{ij})$$

Al realizar las multiplicaciones latentes y al arreglar términos, tenemos:

$$P_{ij} = \gamma_{00} + \gamma_{10}XC_{ij} + \gamma_{20}IHC_{ij} + (u_{0j} + u_{1j}XC_{ij} + e_{ij})$$

Como es usual, los efectos aleatorios se expresan entre el paréntesis, mientras los fijos preceden la secuela de términos de la ecuación mixta. Su interpretación no ha cambiado mucho: γ_{00} representa la gran media de los precios barriales; γ_{10} , la gran pen-

diente de éstos a los ingresos, que expresa el componente común a todas las pendientes ciudadinas; γ_{20} es la gran pendiente del Índice Hedónico centrado, que se considera fija y, por tanto, general para todas las ciudades. u_{0j} es el efecto aleatorio del intercepto, que indica las diferencias entre la gran media y las medias urbanas de los precios barriales: el efecto ciudad sobre las medias. u_{1j} , como se dijo, representa el componente aleatorio de las pendientes.

3. Estimación hipotética e interpretación del modelo 4

El cuadro 1 muestra los efectos fijos, mientras el cuadro 2 exhibe los parámetros de covarianza para el modelo 4. Los tres efectos fijos (γ_{00} , γ_{10} , γ_{20}) son estadísticamente significativos, como lo demuestran sus respectivas pruebas t y los niveles de significación inferiores a 0.05, y no presentan modificaciones sensibles con relación al modelo 3. El primero de ellos nos refleja la gran media de los precios barriales, que equivale a 921. 87 dólares; los otros dos son las gran pendientes del ingreso barrial y del Índice Hedónico, respectivamente. Por cada unidad de diferencia, los precios barriales se distinguen en 0.284, en el caso del primero, y en 0.248, en el del segundo.

Las modificaciones más importantes ocurren en los parámetros de covarianza, que dan cuenta de los efectos aleatorios. En primer lugar, en la notación que acompaña a los resultados de SPSS: UN(1,1) representa la varianza del intercepto o entre las medias ciudadinas con relación a la gran media: σ^2_{u0j} . UN(2,2) es la varianza del segundo efecto aleatorio: el de las pendientes ciudadinas con relación a la gran pendiente, σ^2_{u1j} . Por último, UN(2,1) es la covarianza entre estos dos efectos: el de la media y el de la pendiente. En segundo término, el foco de atención debe de reposar en la varianza del efecto aleatorio de las pendientes (σ^2_{u1j}), que es la indagación principal del modelo 4. Al ser estadísticamente significativo, como lo muestra la prueba Wald Z , informa que es distinto de cero y, por tanto, que las pendientes ciudadinas varían al pasar de una urbe a otra, porque cambia la distancia entre cada una de ellas y la gran pendiente.

Sin embargo, la covarianza UN(2,1) resultó no ser estadísticamente significativa (sig.<0.05). Pero ya se ha dicho que la prueba Wald Z no es del todo confiable y debe sustituirse por las variaciones del estadístico -2LL. En el modelo 3, el valor registrado fue 4,128.7, mientras que en este momento se ana-

liza es 3,898.9. La diferencia entre ambos es 229.8. Por otro lado, el número de grados de libertad es dos, porque ambos modelos difieren en dos parámetros estimados: UN(2,2) y UN(2,1). Esta información conduce a estimar la prueba $\chi^2(2) = 229.8$. El resultado arroja un valor de $p < 0.001 < 0.05$. En este sentido, el parámetro es significativo: existe una asociación entre ambos efectos. Para interpretar esta vinculación de forma más elocuente, Hayes (2002) recomienda transformar la covarianza en un coeficiente de correlación mediante el siguiente procedimiento:

$$r_{01} = \frac{UN(2,1)}{\sqrt{UN(1,1) * UN(2,2)}} = \frac{5.796}{\sqrt{259.902 * 1.683}} = 0.3583$$

De esta forma, es posible afirmar que 35.83% de los cambios en uno de los efectos está asociado positivamente a las variaciones en el otro. Más específicamente: aquellas ciudades cuyos precios responden más al nivel de ingreso barrial, son también las que tienen, en promedio, precios barriales más altos. Y al revés: las de cotizaciones mayores son las que resienten más la influencia del nivel socioeconómico.

VI. Modelo 5: los determinantes del efecto ingreso sobre los precios barriales.

El modelo 4 demostró que existe una diversidad de respuestas de los precios barriales al nivel socioeconómico, aproximado por el ingreso promedio de las colonias (el ingreso barrial). Las siguientes preguntas de la investigación son: ¿por qué existen tales diferencias en las pendientes ciudadinas y ¿cuánto influyen en las cotizaciones promedio de los terrenos coloniales? El modelo 5 se aboca a contestar estos cuestionamientos.

Como se señaló en el primer apartado, la hipótesis central aceptaba matices que se traducían en hipótesis específicas: no sólo el ingreso barrial y los atributos hedónicos de las colonias influyen en los precios medios de los predios; también el mercado puede jugar un papel en sus niveles y en las diferencias que se registran entre ciudades. La variable *proxy* que se eligió para dar cuenta de esta influencia hasta ahora ignorada es la densidad poblacional de

las urbes que albergan a las colonias, como se dijo antes: los kilómetros cuadrados representan, si no cabalmente, sí aproximadamente la oferta disponible de predios, mientras que el número de habitantes acusa la presión de la demanda sobre los mismo. A esta variable se le ha designado como D, y como DC, cuando está centrada por su media.

1. La ecuaciones multinivel del modelo 5: determinantes de las pendientes ciudadinas

En la medida en que la densidad poblacional es un atributo de las ciudades y no de las colonias, esta variable debe ubicarse en el nivel 2; y no, en el primer estrato. La razón estriba en que los usos habitacionales compiten no sólo entre sí sino también con los de índole comercial, público e industrial. La interacción de estos elementos da cuenta de la intensidad de la demanda y de la fragmentación de la oferta. Por tanto, la ecuación para el nivel 1 es la misma que en el modelo 4, mientras la correspondiente al nivel 2 experimenta las siguientes modificaciones:

$$P_{ij} = \beta_{0j} + \beta_{1j}XC_{ij} + \beta_{2j}IHC_{ij} + e_{ij} \quad \text{Nivel 1}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}DC_j + u_{0j}; \beta_{1j} =$$

$$\gamma_{10} + \gamma_{11}DC_j + u_{1j}; \beta_{2j} = \gamma_{20} \quad \text{Nivel 2}$$

- i. La variable DC es introducida sólo con el subíndice "j" para indicar que varía de ciudad a ciudad, y no, entre las colonias de una urbe determinada. A cada una se le asigna este promedio ciudadano.
- ii. La nueva variable incide tanto en el efecto aleatorio de la media de los precios barriales, como en el de la pendiente atribuible a los ingresos, aún cuando sea esta última el foco de interés. El argumento consiste en que una urbe más densamente poblada acarrea una relación oferta-demanda menor que puede traducirse en predios que en promedio son más caros que otra con una más débil densidad poblacional. Así mismo, el coeficiente de correlación entre el efecto aleatorio de las medias y las pendientes reportó que éstas tienden ser más grandes ahí donde aquellas son más abultadas. El corolario es que la densidad también afecta las pendientes ciudadinas.
- iii. Los parámetros que dan cuenta de las medias y

de las pendientes tienen ahora dos efectos fijos cada uno de ellos, en lugar de uno solo, como en los modelos anteriores. Los efectos fijos de las medias son γ_{00} y γ_{01} ; los de las pendientes, γ_{10} y γ_{11} . Este último es el que debe atrapar nuestra atención: en principio, indica en cuánto cambia la pendiente de las ciudades por cada habitante por Km² de diferencia en su densidad poblacional. Pero también es relevante γ_{01} porque reporta cuánta es la variación de las medias por una diferencia similar en aquella densidad.

- iv. Los efectos aleatorios son los mismos que en el modelo 4: además de los residuos, las varianzas de las medias y de las pendientes y la covarianza entre ellos.

2. El modelo mixto de los determinantes de las pendientes ciudadinas.

Si sustituimos las ecuaciones del nivel 2 en las del nivel 1, el resultado es una versión preliminar del modelo mixto:

$$P_{ij} = \gamma_{00} + \gamma_{01}DC_j + u_{0j} +$$

$$(\gamma_{10} + \gamma_{11}DC_j + u_{1j})XC_{ij} + \gamma_{20}IHC_{ij} + e_{ij}$$

Para extraer un mayor provecho de esta versión similar, es conveniente realizar las operaciones latentes y arreglar términos:

$$P_{ij} = \gamma_{00} + \gamma_{01}DC_j + \gamma_{10}XC_{ij} + \gamma_{11}DC_j * XC_{ij} + \gamma_{20}IHC_{ij} + (u_{0j} + u_{1j}XC_{ij} + e_{ij})$$

En esta nueva versión, el modelo mixto presenta una novedad: γ_{11} es el coeficiente que acompaña al producto aritmético de la densidad poblacional y de los ingresos barriales; es decir, dos variables de niveles distintos. Este modelo suele llamarsele “cross level interaction” (modelo de interacción entre niveles). El parámetro reporta en cuánto cambia la respuesta de los precios a los ingresos barriales (la pendiente β_{1j} , que es de nivel 1) cuando dos ciudades difieren en un habitante. Pero vayamos por partes: γ_{00} representa, como siempre, la gran media de los precios barriales. γ_{01} es la pendiente que señala en cuánto se diferencian las medias de estas cotizaciones por cada discrepancia de un habitante/Km2, cuando el promedio de ingresos barriales, por

una parte, y el IH, por el otro, son similares en todas las ciudades. Simboliza, por tanto, la diferencia entre la gran media y la media ciudadina de los precios, cuando aquellas variables asumen un promedio igual todas las urbes. Al someter bajo control a estos predictores, las únicas diferencias entre las ciudades obedecerían a la densidad poblacional. El parámetro γ_{01} da cuenta de esas diferencias en tales circunstancias.

De forma análoga, γ_{10} es la gran pendiente o pendiente media que relaciona los precios con los ingresos barriales, con independencia de la ciudad. Por tanto, γ_{11} marca la diferencia entre las pendientes ciudadinas y esta gran pendiente, cuando nos referimos a la misma relación. Supongamos que una ciudad supera a otra en un habitante por kilómetro cuadrados. El efecto del ingreso en la primera ciudad será $\gamma_{10} + \gamma_{11}$ mientras que en la segunda $\gamma_{10} - \gamma_{11}$. El coeficiente γ_{20} tiene la misma interpretación que en el modelo 4: es la gran pendiente de los precios a IHC; al fijarse, es común para todas las ciudades.

3. Estimación hipotética e interpretación del modelo 5.

El cuadro 1 exhibe la estimación hipotética de los efectos fijos. La media de los precios barriales experimenta un aumento con relación al modelo anterior, precisamente porque este último había soslayado el efecto del mercado. Por su parte, $\gamma_{01} = 0.192$ significa que por cada habitante adicional en la densidad demográfica una ciudad muestra una diferencia de 0.192 dólares con relación a la gran media de los precios barriales. Por ejemplo, si una urbe tiene un habitante más que el promedio, el precio ciudadano promedio de la primera sería igual a $(1,149.216 + 0.192 = 1,149.408)$, si ambas registraran los mismos niveles de ingreso barrial e IH.

Por su parte, el coeficiente de interacción, $\gamma_{11} = 0.035$, nos indica que con ingresos e IH similares, los pendientes ciudadinas que relacionan precios e ingresos se separan de la gran pendiente ($\gamma_{10} = 1.06$) en 0.035 dólares por cada habitante de diferencia en la densidad poblacional. Si una de las urbes tiene un habitante más que el promedio, responderá con $1.06 + 0.035 = 1.095$ dólares de aumento en el precio a un dólar de diferencia en el ingreso barrial.

El cuadro 2 registra los parámetros de covarianza, que dan cuenta de los efectos aleatorios. Todos

resultan significativos, pero en virtud de la complejidad del modelo es conveniente someter las estimaciones a la prueba -2LL. La variación de este estadístico con relación al registrado en el modelo anterior es de 184.8, con una diferencia de dos parámetros estimados. La $\chi^2(2) = 184.8$ arroja un valor $p < 0.001 < 0.05$. Por tanto, las varianzas y la covarianza de estos son distintos de cero.

En la medida en que la densidad es una variable del nivel 2, y, por tanto, un atributo de las ciudades, su introducción debería bajar la varianza del efecto ciudad: u_{0j} . Esta reducción se puede estimar de la siguiente manera:

$$\Delta u_{0j} = 1 - \frac{213.119}{259.902} = 17.97$$

Esto significa que las diferencias interurbanas en la densidad poblacional explica 17.97% de la varianza entre los precios barriales promedio de las ciudades.

Por un procedimiento análogo, se determina la parte explicada de las varianzas en las pendientes:

$$\Delta u_{1j} = 1 - \frac{1.467}{1.683} = 0.1284$$

La densidad demográfica explica 12.84% de la varianza de las pendientes; es decir, de las diferencias entre las respuestas de los precios barriales a la diferencia de ingresos medios ciudadanos.

Conclusiones

El análisis multinivel representa una herramienta útil para el análisis urbano, como se ha pretendido demostrar a lo largo de estas líneas. Nuestro ejemplo hipotético se preguntó inicialmente por que existían diferencias entre los precios promedios de los terrenos de las colonias urbanas, a los cuales llamamos precios barriales. La hipótesis central, aunque simulada, afirmaba que la causa residía principalmente en los diferentes estatus socioeconómicos de las colonias, pero que esa relación estaba matizada por las características hedónicas de las mismas y por la densidad poblacional de las ciudades que las albergan. El ejercicio realizado pretendió ponerse en los zapatos de un investigador que quiere plantearse aquel interrogante y contestarlo mediante técnicas que operativicen las hipótesis. Los distintos modelos

desplegados realizan gradualmente esta tarea. En los primeros la preocupación central reside en las diferencias intra e interurbanas de los precios, fincado en una exploración de interceptos aleatorios; en los dos últimos, la atención la robaron las diferentes sensibilidades ciudadinas de los precios a lo que el investigador consideró su principal determinante: el ingreso urbano barrial, que aproxima el estatus socioeconómico medio de las ciudades. En todas estas etapas, las hipótesis deliberadamente fueron confirmadas porque el su fin era claramente expositivo.

Referencias

- Bonin, Olivier (2010) *Urban location and multi-level hedonic models for the Ile-de-France region*, www.gla.ac.uk/media/media_129762_en.pdf consultado 10/07/2014
- De Leeuw, Jan and Meijer, Erik, (2008) *Handbook of Multilevel Analysis*. New York: Springer
- Gandós, Luis (2014) Stata Avanzado. Análisis Multinivel. Diapositivas. Disponible en <https://es.scribd.com/document/140820249/Analisis-Multinivel-STATA>. Consulta 23/11/14
- Goldstein, Harvey (1999). *Multilevel Statistical Models*. London: Institute of Education, Multilevel Models Projects, April,
- Hayes, Andrew F. (2006). A Primer on Multilevel Modeling. *Human Communication Research* 32 (2006) 385–410
- Hox, Joop (2010). *Multilevel Analysis. Techniques and Applications*. New York: Routledge.
- Luke, D. A. (2004). *Multilevel modeling*. Thousand Oaks, CA: Sage.
- Pardo, Antonio; Ruiz, Miguel Ángel; y San Martín, Rafael (2007). Como ajustar e interpretar modelos multinivel con SPSS, *Psicothema*, año/vol. 19, número 002. pp. 308-301
- Ritchey, F. (2002). *Estadística para las ciencias sociales. El potencial de la imaginación estadística*. México: McGraw-Hill.