

## El Proyecto de Migración Mexicana, Ponderadores<sup>1</sup>

### Introducción

El Proyecto de Migración Mexicana (MMP) recoge datos en lugares de distintos tamaños aplicando encuestas en áreas metropolitanas, ciudades de tamaño medio, pueblos pequeños y ranchos. El MMP obtiene sus muestras de una inmensa geografía. En un principio, el MMP se enfocó en el occidente mexicano (principalmente en los estados de Jalisco, Guanajuato y Michoacán). Posteriormente se extendió para incluir a comunidades localizadas en otros estados, como Oaxaca y Baja California. Los datos del MMP no están explícitamente diseñados para ser representativos de todo México. La muestra es más grande en el occidente mexicano y las comunidades no fueron seleccionadas al azar.

Cuando una ciudad, pueblo o rancho ha sido seleccionado, los directores del proyecto o el supervisor de trabajo de campo delimitan el sitio de estudio, es decir, el área exacta donde se aplicará la encuesta. Esta zona es lo que el proyecto llama la “*comunidad*”. En el caso de una ciudad, se trata usualmente de un barrio o una parte geográficamente específica. En un rancho, tiende a ser el lugar completo. El equipo de campo realiza planos del sitio a encuestar y censa todas las viviendas para crear un marco muestral. Los sitios a ser visitados son seleccionados aleatoriamente de un censo. Normalmente, en cualquier comunidad dada, el equipo de campo del MMP entrevista 200 viviendas *elegibles*. Por ejemplo, es extraordinario que las primeras 200 unidades obtenidas del censo puedan ser elegibles, ya que, por ejemplo, son *no elegibles* para la encuesta casas vacías y locales comerciales dónde nadie reside. El proyecto estima el número de casas elegibles en el sitio de la encuesta como el número total de viviendas censadas multiplicada por la proporción de hogares elegibles entre los hogares seleccionados aleatoriamente del marco muestral.

Para cada comunidad seleccionada en México, el MMP intenta una muestra de inmigrantes establecidos Estados Unidos. Esto no siempre es posible, sobre todo, en el caso de comunidades mexicanas con pocos trabajadores migrantes residentes en aquel país. En la mayoría de casos, sin embargo, el proyecto ha completado las muestras mexicanas con su correspondiente muestra en Estados Unidos — las muestras son 57 de las 71 comunidades en MMP71. El equipo de campo del MMP en Estados Unidos tiene notas del campo y referencias para localizar inmigrantes de comunidades específicas establecidas al norte de la frontera. Aquí, el procedimiento de selección no involucra censo ni selección aleatoria de viviendas. Se usa un cálculo especial para estimar el tamaño de la población inmigrante.

---

<sup>1</sup> Estas notas aplican a los ponderadores del Proyecto de Migración Latinoamericana

## Cálculo de ponderadores

En investigación por encuestas basadas en muestras de grupos específicos (como minorías, desempleados o militares) los ponderadores son esenciales para el cálculo correcto de la estadística descriptiva de la población de estudio. El fallo en su aplicación llevaría a las declaraciones simplistas, como que el 50% de la población está desempleada o que se enroló en el ejército. El MMP no hace intencionalmente muestras para grupos específicos, sino para hogares dentro de las comunidades que son seleccionados aleatoriamente. Aquí, los ponderadores del MMP no varían por hogar o individuo, en lugar de ello, varían por comunidad y muestra específica.

Hay dos ponderadores por comunidad: uno para la muestra mexicana y otro para la muestra de EE.UU. Una muestra ponderada se calcula como el inverso de la *fracción muestral*. En el caso de los ponderadores para las comunidades mexicanas (variable *mxweight*)<sup>2</sup>, la fracción muestral es obtenida por la división del número de hogares entrevistados entre la estimación de los hogares elegibles en el marco muestral. El cuadro 1 muestra el cálculo de los ponderadores para la comunidad mexicana número 58. La información necesaria para estos cálculos es suministrada por la plantilla de registro del supervisor de campo (pida una copia del Manual del Supervisor de Trabajo de Campo del MMP y LAMP para más detalles)

**Cuadro 1. Cálculo de ponderador en la muestra mexicana, comunidad 58, MMP71**

Total de hogares en el lugar de la muestra (conteo)	254
Estimación de hogares elegibles (1)	243
Hogares visitados	118
elegibles	113
<b>Hogares entrevistados</b>	<b>100</b>
Descartados (2)	10
Rechazos	3
<b>No elegibles</b>	<b>5</b>
Casas vacías (3)	3
Jefe extranjero	0
Negocio	2
Inexistente (4)	0
En EEUU (5)	0
Otros inelegible (6)	0
Fracción muestral (7)	0.41
Ponderador (8)	2.43

(1) hogares contados \* (elegibles visitados / total visitados)

<sup>2</sup> Variable *doweighth*, en las bases de datos del LAMP

- hogares habitados cuyos residentes nunca pudieron ser encontrados
- (2) o no eran capaces de responder la encuesta de manera coherente.
  - (3) casa vacía o en construcción y aún no habitada. considerado por error un hogar independiente al momento del conteo
  - (4) de hogares. hogares sin residentes cuyos dueños, según los vecinos, se
  - (5) encontraban "en los Estados Unidos". instituciones tales como escuelas, hogares de ancianos, oficinas del
  - (6) gobierno, etc.
  - (7) hogares entrevistados / estimación de hogares elegibles
  - (8) 1 / fracción muestral

El ponderador para Estados Unidos, al igual que el mexicano, se calcula como el inverso de la fracción muestral, el cual resulta de dividir el número de hogares entrevistados en EE.UU., por el tamaño total (número de hogares) de la comunidad americana. Ya que las muestras americanas son construidas a través de las referencias (un estudio por muestreo es financieramente inviable), la población deberá ser estimada usando métodos indirectos. El método de estimaciones selectas de la población en comunidades norteamericanas es una proporción de la población de comunidades mexicanas bajo estudio. La pregunta principal es: ¿cómo hacer el número de hogares en una comunidad x americana (hogares cuyos jefes son originarios de una comunidad x en México) comparable al número de hogares en comunidad x (de la población mexicana)?

Para responder esta pregunta, el proyecto usa datos de los hijos de los jefes de hogar en la muestra mexicana. El cuestionario del MMP recoge datos de todos los residentes del hogar, incluyendo a los hijos del jefe que no son miembros del hogar. Obtenemos la proporción que necesitamos para así comparar el número de hijos establecidos en Estados Unidos con aquéllos que dejaron hogar paterno pero que siguen en México. Aplicando esta proporción a la población mexicana (la estimación de hogares elegibles dentro del marco muestral) estimamos la población americana (número total de casas americanas) para cada comunidad. El cuadro 2 ejemplifica el cálculo de los ponderadores americanos para la misma comunidad del cuadro 1.

**Cuadro 2: Cálculo de ponderadores en la muestra de Estados Unidos, Comunidad No. 58 del MMP71**

a) Número total de hijos del hogar de la muestra de origen, quienes no viven en el hogar y aquellos que en uscurtrp son desconocidos <i>Surveypl=1, relhead=3, hhmemshp=1 y uscurtrp ni 9999</i>	165
<b>De los seleccionados en (a) ...</b>	

b) Número total de hijos viviendo en EE.UU. <i>En uscurtrp=1</i>	13
c) Número total de hijos migrantes viviendo en el país de origen <i>En uscurtrp=2</i>	0
d) Número total de hijos no migrantes viviendo en el país de origen <i>En uscurtrp=8888</i>	152
e) Número total de hijos viviendo en el país de origen <i>c+d</i>	152
f) Promedio de hijos en EE.UU. sobre los hijos del país de origen <i>b/e</i>	0.086
g) Ponderadores de la muestra de las comunidades de origen <i>De la hoja de registro del supervisor de campo</i>	2.43
h) Tamaño muestral de la comunidad de origen <i>N de la comunidad</i>	100
i) Hogares elegibles estimados en la comunidad de origen <i>g*h</i>	243.237
j) Número estimado de hogares en EE.UU <i>f*i</i>	20.803
<b>k) Número total de hogares entrevistados en EE.UU.</b> <i>N para la muestra de EE.UU</i>	10
m) Fracción muestral de EE.UU estimada <i>k/j</i>	0.4807
n) Ponderador de la muestra de EE.UU: inverso de la fracción muestral <i>1 / m</i>	2.080

### El significado actual de los ponderadores del MMP

Cuando se emplean, los ponderadores para México producen datos representativos del área formada por todo el marco muestral. Esta área no es “México”, ni tampoco “El occidente de México”, no es la combinación de estados incluidos en el estudio. No es la combinación de ciudades, pueblos y ranchos donde se aplicó la encuesta. Es la combinación de de la población de todas las comunidades muestreadas. Actualmente, el área de la muestra es bastante grande, incluye 114 comunidades esparcidas a lo largo de mexicano y su contraparte en Estados Unidos. Así como no es geográficamente un área continua, tampoco es un área diseñada en el tiempo. Cada comunidad de MMP fue inspeccionada una vez, entre 1982 y 2006.

## El Análisis descriptivo

El uso de ponderadores se recomienda para el cálculo de estadísticas descriptivas con propósitos analíticos. Por otra parte, estas estadísticas podrían ser parciales en poblaciones muestradas en pueblos o ranchos donde la fracción muestral tiende a ser grande. De la misma manera, sin el ponderador, los hogares en áreas urbanas tendrían baja representación ya que la probabilidad de ser entrevistados es generalmente bajo que en un pueblo pequeño o rancho.

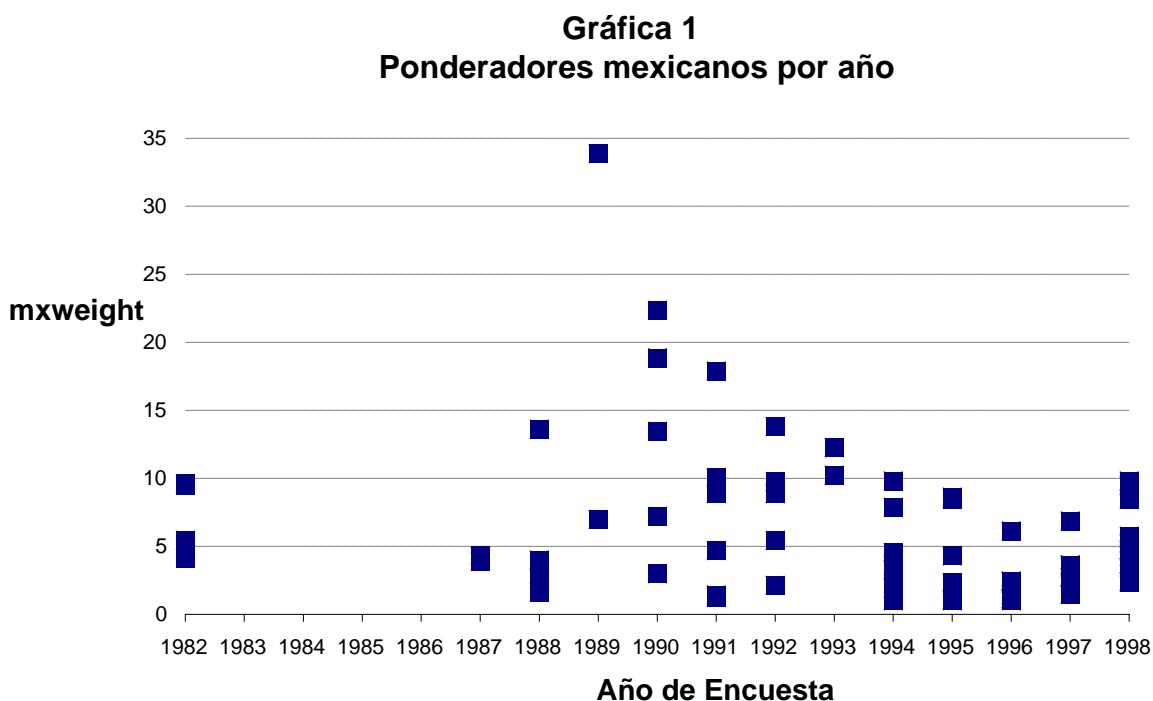
Por ejemplo, considere las comunidades 37, 44 y 45 del MMP, los cuales son ranchos y muestran el valor más bajo del ponderador *mxweight* (1.00) y las comunidades 9 y 11, áreas urbanas que presentan el valor más grande del ponderador *mxweight* as (33.88 y 22.36, respectivamente). El Cuadro 3 muestra el porcentaje de trabajadores migrantes masculinos y el promedio de viajes a EE.UU., para hombres mayores de 15 años, sin ponderar y ponderados. Puesto que la migración es más alta en las comunidades más pequeñas, las estadísticas sin ponderar sobrestiman porcentaje de trabajadores migrantes y el promedio de viajes a EE.UU. Para las cinco comunidades analizadas, las áreas más pequeñas tienen mayor probabilidad de ser completamente muestradas y mayor probabilidad de tener una alta fracción muestral. Lo cual produce pequeños ponderadores.

**Cuadro 3: Hombres mayores de 15 años, muestra mexicana, MMP71, comunidades 9, 11, 37, 44, 45. Medias de variables seleccionadas sin ponderar y ponderadas**

Estadístico	Comunidad					Todas combinadas	
	9	11	37	44	45	Sin ponderar	Ponderadas
% de migrantes	44.9	38.7	78.9	53.5	54.3	52.0	43.3
Media de USTRIPS	1.51	1.17	2.37	2.03	2.26	1.81	1.42
MXWEIGHT	33.88	22.36	1.00	1.00	1.00		
N	492	424	289	275	416		

Este simple ejemplo es suficiente para ilustrar los ponderadores en estadística descriptiva con el MMP. Aun hay algunas preguntas acerca de la delimitación arbitraria de los sitios de estudio. El denominador en el cálculo de los ponderadores es la población total de hogares, que depende de las dimensiones del sitio del estudio. Los equipos de campo más grandes tienen mayor probabilidad de hacer mapas más completos de las comunidades que los equipos más pequeños. En este sentido, los ponderadores tienen alguna dependencia de los recursos del Proyecto en el momento en que se realiza el estudio de campo. Más allá de esta variabilidad inevitable, el MMP hace a un esfuerzo por mantener procedimientos de campo consistentes y evitar demasiada variabilidad en los tamaños de los sitios a encuestar. La

gráfica 1 muestra que, después de que el proyecto regresó al campo en 1987, estos esfuerzos han mejorado con el tiempo, manteniendo relativamente niveles moderados de variabilidad.



Algunas veces, las estadísticas descriptivas se reportan con propósitos analíticos. La situación típica consiste en una tabla de medias y desviación estándar que preceden a un análisis de regresión. En este caso, si la regresión procede de datos sin ponderar, los datos no serán razonables cuando se reporten medias y desviaciones estándar con datos ponderados. Los lectores probablemente estarán más interesados en las medias y las desviaciones estándar de las variables incluidas en un modelo causal tal cómo estas son – así estas serán usadas en el modelo.

Los ponderadores americanos compensan las variaciones de los tamaños estimados de las comunidades en estados Unidos y proveen una herramienta para la medición de la contribución relativa de cada muestra de comunidades migrantes binacionales. Los ponderadores las la muestra de EE.UU. no resuelven el problema de de no representatividad de la muestra en ese país, la cual se construyó a través de referencias usando el método de “bola de nieve”. El investigador puede omitir totalmente las muestras de EE.UU. por razones estadísticas. Dependiendo del problema de investigación, esto puede o no limitar el alcance del análisis.

## **Modelando análisis causal**

Los datos del MMP son muy convenientes para la estimación de modelos causales que intentan verter luz en las complejidades del proceso de migración. Massey y Zenteno (2000; Zenteno y Massey 1999, en español), mediante una muestra parcial, señalan que los datos del MMP son aproximadamente comparables a la muestra representativa de la Encuesta Nacional de la Dinámica Demográfica (ENADID). Incluso, los datos de MMP son superiores a la ENADID al estimar modelos causales, por dos razones. Primero, la inmensa cantidad de información cuantitativa contenida en el banco de datos de MMP permite modelos complejos. Segundo, al incluir datos de los hogares emigrantes localizados en Estados Unidos, el MMP reduce problemas de selectividad encontrados en la ENADID. Esta última sólo inspecciona hogares en México.

Tal como declara Massey et al. (1987: 12-13): *La etnoencuesta no es, por supuesto, la última palabra en estudios de migración internacional. Aun se enfrenta con el problema de generalización. La etnoencuesta no es una técnica para la estimación estadística. Los cuadros y gráficas contruidos con los datos de la etnoencuesta no pueden ser extrapolados fácilmente al resto de México o a la población de trabajadores emigrantes mexicanos. Lo que el método hace es proporcionar una manera de comprender e interpretar los procesos sociales que están debajo del conjunto de las estadísticas. La fuerza de la etnoencuesta está en proporcionar datos duros para que el proceso social de la migración internacional pueda describirse de manera contundente y convincente.*

Cuando en los datos del MMP son usados (para estimar modelos causales) ponderadores, en la mayoría de los casos es innecesario. Los modelos de la regresión asumen que los coeficientes son los mismos para todos en la muestra. Si ese es el caso, no hay ninguna justificación para ponderar. Si no es así, probablemente la solución es re-especificar el modelo, por ejemplo incluyendo los términos de interacción o segmentar modelos separados para grupos de población diferentes.

Winship y Radbill (1994) muestran que cuando el modelo se ha especificado correctamente y los ponderadores son solamente una función de las variable independientes, entonces los estimados de regresión logística sin ponderar serán imparciales y consistentes y tendrán pequeños errores estándar que los estimados de regresión logística con ponderador. Así, el uso de muestras ponderadas producirá parámetros estimados menos eficientes que al usar muestras sin ponderar (Cuadro1).

Además, esto demuestra que las sensibles diferencias entre las estimaciones derivadas del uso de muestras ponderadas y sin ponderar son buenos indicadores de que el modelo no

está correctamente especificado o que los ponderadores son una función de la variable dependiente. Cuando esto último es el caso, los autores recomiendan reconsiderar el modelo especificado. Si no es posible re-especificar el modelo, entonces el uso de muestras ponderadas puede ser más apropiado acompañado de estimadores heteroscedásticos consistentes de White para los errores estándar (White, 1980).

## Referencias

Massey, Douglas S., Rafael Alarcón, Jorge Durand, and Humberto González. 1987. *Return to Aztlan: The Social Process of International Migration from Western Mexico*. Berkeley and Los Angeles: University of California Press.

Massey, Douglas S. and René Zenteno. 2000. "A Validation of the Ethnosurvey: The Case of Mexico-U.S. Migration." *International Migration Review* 34(3):766-793.

Winship, Christopher and Larry Radbill. 1994. "Sampling Weights and Regression Analysis." *Sociological Methods and Research* 23(2):230-257.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity". *Econometrica* 48:817-838.

Zenteno, René M. and Douglas S. Massey. 1999. "Especificidad versus representatividad: enfoques metodológicos en el estudio de la migración mexicana hacia Estados Unidos." *Estudios Demográficos y Urbanos*: 14(1):75-116.