

Ponderadores en el proyecto sobre migración Mexicana¹

Introducción

El proyecto sobre migración mexicana (MMP) recolecta datos en lugares con características varias como son las áreas metropolitanas, ciudades pequeñas, pueblos y ranchos. Las comunidades muestreadas en el MMP se encuentran geográficamente esparcidas en México. Es cierto que el MP en sus inicios se enfocó en el occidente de México, principalmente en los estados de Jalisco, Guanajuato, y Michoacán. Pero hasta la fecha, el MMP se ha expandido para incluir comunidades en otros estados de México como son Yucatán ó Veracruz. El MMP no está diseñado explícitamente para representar a México como una unidad, sino que el MMP toma muestras de varias comunidades no seleccionadas al azar.

Una vez que una ciudad, pueblo ó rancho han sido seleccionados, los directores del proyecto ó el supervisor de campo delimitan el área a encuestar, es decir, el área precisa en dónde las encuestas serán llevadas a cabo. A esta delimitación el proyecto llama “comunidad.” En una ciudad, el área delimitada es usualmente un vecindario ó una parte geográficamente específica de éste. En un rancho, la comunidad suele ser todo el lugar. El equipo de trabajo de campo mapea todo el sitio y enumera todos los hogares dentro de los límites especificados para crear la *estadística muestral*. Los hogares a ser entrevistados son elegidos al azar del conteo total. En todas las comunidades, el MMP típicamente encuesta 200 hogares *elegibles*. Es muy raro que los primeros 200 hogares elegidos al azar sean elegibles; por ejemplo, casas vacías o negocios locales no habitables se consideran *no elegibles*. El proyecto calcula el número total de hogares elegibles en el área delimitada como el número total de hogares enumerados por la proporción de hogares elegibles entre los hogares que fueron seleccionados al azar en el área delimitada.

Para cada comunidad encuestada en México, el MMP trata de encuestar emigrantes de esa misma comunidad que viven en los Estados Unidos. Esto no es siempre posible, especialmente en casos dónde solamente hay unos cuantos emigrantes residiendo en los Estados Unidos. En muchos casos, el proyecto ha podido complementar la muestra mexicana con la muestra

¹ Estas notas también aplican a los ponderadores para el proyecto sobre migración Latino Americana (LAMP).

correspondiente en los Estados Unidos. Hasta el 2012, el MMP ha recolectado 68 muestras en Estados Unidos de las 134 comunidades que se encuentran en el MMP¹³⁴. Los trabajadores de campo del MMP en los Estados Unidos trabajan con las notas de trabajo de campo que se recabaron en México y con recomendaciones para localizar a emigrantes de comunidades específicas que viven en el norte. Por ello, el procedimiento para el ponderador de la muestra estadounidense no involucra ni una enumeración de hogares ni una muestra al azar. Para estimar el tamaño de la población emigrante son necesarios una serie de cálculos especiales los cuales se describen a continuación.

¿Cómo se calculan los ponderadores?

En lugares donde se encuestan de más a grupos específicos de población (ya sean minorías, desempleados, ó militares), los ponderadores son esenciales para el computo correcto de las estadísticas descriptivas de la población que se estudia. Si no se aplican los ponderadores, se podría llegar a falsas conclusiones como, por ejemplo, que el 50% de la población está desempleada ó están en la milicia. El MMP no se enfoca en un grupo de población en específico, ya que los hogares son elegidos al azar. Entonces, los ponderadores del MMP no varían por hogares ni por individuos. Al contrario, los ponderadores que ofrece el MMP son específicos a cada comunidad y a cada muestra.

Hay dos tipos de ponderadores para cada comunidad: uno es el ponderador para la muestra de México y el otro es el ponderador para la muestra de Estados Unidos. Un ponderador es calculado como el inverso de la fracción muestral. En el caso de los ponderadores de la muestra mexicana (variable *mxweight*)², la fracción muestra se obtiene al dividir el número de hogares entrevistados entre el número de hogares elegidos en la muestra. La tabla 1 muestra los cálculos para el ponderador de la comunidad 106. Existen más detalles ya documentados y se pueden encontrar en la guía del supervisor del trabajo de campo que se encuentra en la página del MMP.

² Para las muestras del LAMP, la variable es *doweight*.

Tabla 1: Cálculo del ponderador, MMP134, Comunidad #106

Estadísticas básicas	
total de hogares en el lugar de la muestra (conteo)	1483
estimación de hogares elegibles (1)	1114
hogares visitados	293
elegibles	220
hogares entrevistados	201
descartados (2)	3
rechazos	16
no elegibles	73
casas vacías (3)	28
jefe extranjero	0
negocios	7
inexistente (4)	7
en EEUU (5)	31
otros inelegibles (6)	0
Fracción muestral (7)	0.1804
Ponderador (8)	5.5423
Taza de rechazo	0.074

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

El ponderador de la muestra de Estados Unidos, al igual que el ponderador de la muestra mexicana, se calcula como el inverso de la fracción muestral. La fracción muestral es el resultado de dividir el número de hogares entrevistados en los Estados Unidos entre el tamaño total de hogares de dicha comunidad en los Estados Unidos. Debido a que las muestras en Estados Unidos se construyen a través de referencias, la población se debe *estimar* utilizando métodos indirectos. El método que el MMP utiliza para estimar la población de dicha comunidad en los Estados Unidos es la proporción de la población mexicana de dicha comunidad bajo estudio. Entonces la pregunta es: ¿cómo se compara el número de hogares de la comunidad X en Estados Unidos con el número de hogares de esa misma comunidad X en México?

Para responder la pregunta anterior, el proyecto utiliza los datos referentes a los hijos de los jefes de hogar en la muestra mexicana. La etnoencuesta del MMP recolecta información de todos los residentes del hogar que también incluye a los hijos del jefe del hogar que ya no son considerados miembros del hogar al momento de la encuesta. Con esto, podemos obtener la proporción que necesitamos para comparar el número de hijos que viven en los Estados Unidos con aquellos hijos que han dejado el nicho familiar pero que se han quedado en México. Al aplicar la dicha proporción a la población mexicana (hogares elegibles dentro de la muestra), podemos estimar la población en los Estados Unidos (número total de hogares en EE.UU) para cada comunidad. La tabla 2 muestra un ejemplo de cómo calculamos el ponderador para la muestra en Estados Unidos para la misma comunidad que se mostró en la tabla 1.

Tabla 2: Cálculo del ponderador de la muestra en EE.UU., MMP134, Comunidad #106

a) Número total de hijos (en la muestra mexicana) que no viven en el hogar y para quienes el viaje actual no es desconocido surveypl=1, relhead=3, hhemshp=1 and uscurtrp ne 9999	241
De los seleccionados en (a)...	
b) Número total de hijos viviendo actualmente en los EE.UU. surveypl=1, relhead=3, hhemshp=1 and uscurtrp=1	10
c) Número total de hijos migrantes que viven en México surveypl=1, relhead=3, hhemshp=1 and uscurtrp=2	2
d) Número total de hijos no migrantes que viven en México surveypl=1, relhead=3, hhemshp=1 and uscurtrp=8888	229
e) Número total de hijos viviendo en México c + d	231
f) Proporción de hijos en los EE.UU. a los hijos que viven en México b / e	0.043
g) Ponderador de la muestra doméstica Se obtiene de las estadísticas muestrales	5.54
h) Tamaño de la muestra en el país de origen N para esta comunidad	201
i) Número estimado de hogares elegibles en la comunidad de origen g * h	1114
j) Número estimado de hogares en los EE.UU. f * i	48.21
k) Número total de hogares entrevistados en los EE.UU. N para la muestra de los EE.UU.	21
m) Fracción muestral estimada de los EE.UU. k / j	0.44
n) Ponderador de la muestra en los EE.UU. 1 / m (inverso de la fracción muestral)	2.30

El significado actual de los ponderadores del MMP

Cuando los ponderadores del MMP son utilizados, éstos producen datos representativos del área conformada por la muestra. Esta área no es “México” ni el “Occidente de México” ni la combinación de estados incluidos en la etnoencuesta. Tampoco es la combinación de ciudades, pueblos ó ranchos dónde se ha realizado el trabajo de campo. Los ponderadores *son* la combinación de la población de todas las comunidades entrevistadas. Hasta el 2012, el área es grande ya que incluye 134 comunidades esparcidas por todo México y por mucho territorio en los EE.UU. Debido a que no entrevistamos un área geográficamente continua, los ponderadores tampoco representan un área en un tiempo específico. Cada comunidad del MMP ha sido entrevistada solo una vez entre 1982 y el 2012.

Análisis descriptivos

El uso de ponderadores se recomienda cuando se trabaja con estadísticas descriptivas con propósitos analíticos. De otra manera, estas estadísticas no serían imparciales en la dirección determinadas por las características de la población entrevistada en pueblos y ranchos en dónde las fracciones muestrales tienden a ser más altas. Por lo contrario, sin ponderadores, los hogares en las áreas urbanas estarían subrepresentados ya que la probabilidad de ser entrevistados es generalmente menor a los hogares localizados en pueblos ó ranchos (vea la figura 1 y su explicación).

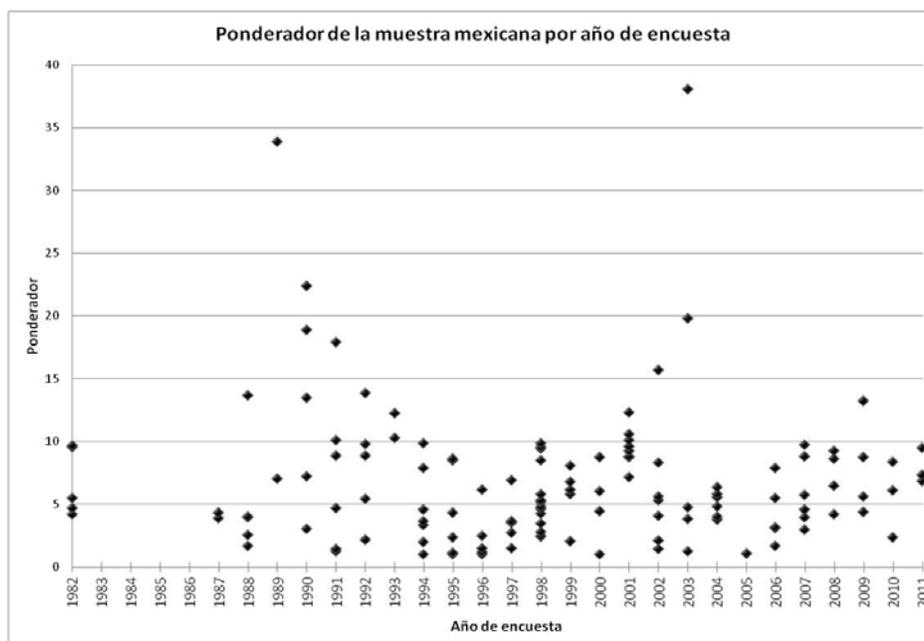
Por ejemplo, consideremos las comunidades 37, 44, y 45 que son ranchos y presentan un ponderador muy pequeño en la muestra nacional de 1. También, consideremos las comunidades 9 y 11, que son comunidades localizadas en áreas urbanas con ponderadores muy altos en la muestra nacional – 33.88 y 22.36 respectivamente. Si seleccionamos a hombres mayores de 15 años, la tabla 3 muestra el porcentaje de hombres migrantes y el promedio de viajes a los EE.UU. con y sin ponderador. Ya que la migración es más prevalente en pequeñas comunidades, las estadísticas sin ponderador sobreestiman ambos números: el porcentaje de migrantes y el número promedio de viajes a los EE.UU. Para las cinco comunidades analizadas, las áreas más chicas tienen la probabilidad de ser encuestadas en su totalidad y por ello muestran fracciones muestrales altas, lo que produce ponderadores muy pequeños.

Tabla 3: Hombre mayores de 15 años, de las muestras mexicanas del MMP134, comunidades 9, 11, 27, 44, 45: Promedios con/sin ponderador para variables seleccionadas

Estadística	Comunidad					Todas juntas	
	9	11	37	44	45	Sin ponderador	Con ponderador
% migrantes	44.9	38.7	78.9	53.5	54.3	52.0	43.3
Promedio de viajes a los EE.UU.	1.51	1.17	2.37	2.03	2.26	1.81	1.42
mxweight	33.88	22.36	1.0	1.0	1.0		
N	492	424	289	275	416		

Este simple ejemplo puede ser suficiente para las estadísticas descriptivas del MMP sean ponderadas. Pero, todavía hay preguntas del hecho de cómo se determinan arbitrariamente los límites del lugar a entrevistar. Para calcular los ponderadores, el denominador es la población total de hogares, el cual depende de la dimensión del área a entrevistar. Si el equipo de trabajo de campo es grande, el supervisor puede delimitar un área más grande; pero si el equipo de trabajo es más pequeño, el área a delimitar será más pequeña. Sabiendo esto, el MMP trata de mantener consistente el tamaño del equipo de trabajo de campo año con año para evitar variación innecesaria como la que acabamos de mencionar; generalmente, el equipo de trabajo de campo del MMP consta de un supervisor y cuatro encuestadores. La gráfica 1 muestra que desde 1987, los esfuerzos del proyecto en ser consistente han mejorado a través del tiempo, ya que la variabilidad en los ponderadores se ha mantenido moderada.

Gráfica 1



Algunas veces, las estadísticas descriptivas son utilizadas para *reportar* y no para *propósitos analíticos*. En un análisis típico, antes de presentar los resultados de la regresión, primero se presenta una tabla de promedios y desviaciones estándar. En el caso que la regresión no pondere los datos, es razonable que tampoco se ponderen los promedios ni las desviaciones estándar de la muestra. Los lectores seguramente estarán más interesados en los promedios y desviaciones estándar que están incluidas en el modelo de la misma manera que serán usadas en la regresión.

Los ponderadores de la muestra de EE.UU. compensan por los tamaños estimados de las comunidades migrantes en los EE.UU., y proveen una herramienta de medida de la contribución relativa de cada muestra hecha en los EE.UU. a la muestra binacional de cada comunidad. Debido a que las muestras de EE.UU. se construyen a través de referencias utilizando el método de la “bola de nieve,” el ponderador asignado a cada muestra no resuelve el problema de representatividad. El investigador puede elegir el no utilizar estas muestras en su análisis. Dependiendo del análisis, esto limitar o no el rango de análisis.

Modelos Causales

Los datos del MMP son los más adecuados para estimar modelos causales que investigan los procesos complejos que involucra el proceso migratorio. Massey y Zenteno (2000, 1999) demuestran que los datos del MMP, a pesar de que son elegidos al azar, son muy comparables a los que tiene la Encuesta Nacional de la Dinámica Demográfica (ENADID), la cual es una muestra representativa a nivel nacional. De hecho, los datos del MMP son superiores a los de la ENADID cuando se usan para el análisis de modelos causales por dos razones: primera, la vasta información cuantitativa que el MMP tiene permite modelos complejos; y segunda, al incluir los hogares localizados en los Estados Unidos del MMP, los problemas de selectividad se reducen, ya que la ENADID solo hace encuestas en México.

Massey *et. al.* (1990:22) escribieron:

“Por cierto, la etnoestadística no es la última palabra en el estudio de la migración internacional, ya que todavía se enfrenta con el problema de la generalización. La etnoestadística no es una técnica para totalizar una estimación estadística. Hechos y números calculados con base en la información etnoestadística no pueden fácilmente

generalizarse al resto de México, o a la población de emigrantes mexicanos. Lo que el método proporciona es una forma de entender e interpretar los procesos sociales que se fundamentan en las estadísticas reunidas. La validez de la etnoestadística es que provee información consistente de modo que el proceso social de la migración internacional puede describirse de una manera fundamentada y convincente.”

En la mayoría de los casos, cuando los datos del MMP son usados en modelos causales, el utilizar los ponderadores es innecesario. Los modelos de regresión asumen que los coeficientes son los mismos para todos en la muestra. Si eso es verdad, entonces no hay una justificación para utilizar ponderadores. Si eso no es verdad, la solución más sensata sería revalorar el modelo, por ejemplo, incluir interacciones apropiadas o hacer modelos estadísticos para diferentes grupos de población.

Winship y Radbill (1994) demuestran que cuando un modelo ha sido especificado correctamente y los ponderadores son una función independiente de las variables, entonces los estimados de la regresión logística sin ponderador son imparciales, consistentes, y tienen errores estándares pequeños que los estimados de la regresión logística con ponderador. Entonces, el uso de ponderadores producen parámetros menos eficientes que aquellos sin ponderar (ver tabla 1 y p. 244). Además, ellos demuestran que si existen diferencias notables entre los estimados que resultan de una muestra ponderada y una sin ponderar, entonces hay una buena indicación que el modelo no fue especificado correctamente, ó puede ser que los ponderadores son una función dependiente de la variable. Cuando los ponderadores son una función dependiente de la variable, los autores recomiendan el reconsiderar la especificación del modelo. Si esto no fuera posible, entonces el uso de ponderadores puede ser más apropiado, junto con los estimadores heteroscedasticos de errores estándar de White (White 1980). En el caso de los datos del MMP, los ponderadores son una función de las variables de comunidad y lugar de encuesta, y no hay necesidad de utilizar ponderadores para modelos causales.

Referencias:

Massey, Douglas S., Rafael Alarcón, Jorge Durand, and Humberto González. 1990. *Los Ausentes: El Proceso Social de Migración Internacional en México Occidental*. Colección Los Noventa, no. 61. México, D.F.: Alianza Editorial Mexicana, Consejo Nacional para la Cultura y las Artes.

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

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