



Fondo de la OIM para el Desarrollo
FOMENTO DE CAPACIDADES EN GESTIÓN DE LA MIGRACIÓN



Centro Centroamericano
de Población

Diseño y selección de una muestra para la población migrante de Costa Rica

María Fernanda Alvarado Leitón
Centro Centroamericano de Población

-Noviembre 2015-

1. Introducción

Desde el año 2013 se establece una relación entre el Centro Centroamericano de Población (CCP) y la Organización Internacional para las Migraciones (OIM) con el fin de contribuir a la generación de información y estadísticas confiables sobre migración en Costa Rica, que pueda ser tomada en cuenta en las agendas gubernamentales e institucionales del país.

De ahí nace el proyecto CCP-OIM PRO129 ejecutado desde el CCP cuya finalidad es realizar el levantamiento de datos que permita cumplir con el objetivo antes descrito. Se plantea entonces la Encuesta Nacional de Migración, una encuesta de hogares con el propósito de captar el estado actual y dinámicas de la población inmigrante y emigrante.

Para la decisión de las temáticas a captar en el cuestionario se realizaron reuniones con las instituciones públicas y organizaciones no gubernamentales involucradas en las políticas relacionadas a la migración y el trato a los migrantes. Otras decisiones administrativas, como los recursos disponibles y requerimientos de la muestra han sido discutidos internamente entre el CCP y la OIM.

Con base en las temáticas discutidas en las reuniones con las instituciones, se plantea un cuestionario que pretende captar las dinámicas migratorias de los inmigrantes y emigrantes en Costa Rica. Este cuestionario se presentó nuevamente a las instituciones y se abrió un período de discusión en el que las instituciones pudieron enviar sus comentarios y sugerencias sobre preguntas específicas.

El presente trabajo se enmarca dentro de este proyecto como el siguiente paso posterior a la realización del cuestionario y tiene como objetivo describir los detalles técnicos y operativos del diseño muestral a utilizar. Se presenta en él las decisiones tomadas en relación al tamaño de muestra, el marco muestral y el diseño muestral propiamente. Además se incluyen las características de las poblaciones de estudio que llevan a las decisiones tomadas y una simulación de la muestra realizada utilizando el marco muestral.

2. Objetivos

El objetivo principal del presente trabajo es generar un diseño de muestra probabilístico, que permita obtener datos que respondan a las preguntas de investigación, esté acorde al presupuesto disponible y no presente complicaciones excesivas de ejecución.

Por otra parte, el objetivo general de la encuesta es la medición de los cambios en las condiciones socioeconómicas de las personas (condición laboral, seguridad social, nivel educativo, vivienda, etc.) posterior a la experiencia migratoria.

Por lo tanto, en su mayoría se estimarán proporciones para presentar los resultados de la encuesta. Asimismo, se desea obtener algunos totales a nivel nacional con el único fin de validar la encuesta con cifras existentes.

3. Población de estudio

Para esta encuesta, se tienen dos poblaciones de interés: los emigrantes y los inmigrantes. Se consideran como **inmigrantes** personas de 15 años o más que tienen al menos un mes de residir en Costa Rica al momento de la encuesta y planean quedarse al menos 3 meses para vivir y/o trabajar. Asimismo, son **emigrantes** aquellas personas de 15 años o más que tienen al menos un mes de haber salido de Costa Rica al momento de la encuesta y planean quedarse al menos 3 meses en el país de destino.

4. Construcción del Marco Muestral

En el apartado anterior se describieron las poblaciones de interés para la encuesta. En consecuencia, se plantearon inicialmente dos marcos muestrales distintos para dos muestras distintas y se evaluó posteriormente su traslape ya que no se considera adecuado asumir que ambas poblaciones son completamente independientes y se encuentran en zonas geográficas mutuamente excluyentes.

La confección de estos marcos muestrales se llevó a cabo utilizando la información del X Censo Nacional de Población y VI de Vivienda realizado en 2011 por el INEC. Debido a que las viviendas en las que residen los emigrantes y los inmigrantes no son identificables en el censo, los marcos muestrales se conforman de áreas geográficas; en este caso, de unidades primarias de muestreo (UPM). Sin embargo, dado que la información a nivel de UPM no se encuentra disponible en el censo, algunas decisiones iniciales del diseño se basan en la información disponible a nivel distrital.

Como primera aproximación al diseño de la muestra se calcularon algunas cifras con base en el censo con el fin de describir las características generales de las poblaciones de estudio (cuadro 1). Se encontró que según lo reportado, para el año 2011 en Costa Rica residían en hogares particulares 382 251 personas nacidas en el extranjero; asimismo se obtuvo que 59 217 personas que residían en hogares particulares dejaron su hogar para vivir en otro país.

En términos de hogares, lo anterior se traduce en 199 329 (hogares con al menos 1 inmigrante 16,11% del total nacional de hogares), 43 240 hogares con al menos 1 emigrante (3,50% del total nacional de hogares) y 13 046 hogares en los que se encontraron tanto emigrantes como inmigrantes (1,05% del total nacional de hogares).

Cuadro 1
Cifras generales sobre la población emigrante e inmigrante de Costa Rica

Característica	Cifra
Número de personas en CR	4 301 712
Número de personas en CR en hogares particulares	4 283 063
Número de hogares particulares en CR	1 236 981
Número de inmigrantes en hogares particulares	382 251
Número de emigrantes que residían en hogares particulares	59 217
Número de hogares particulares con al menos 1 inmigrante	199 329
Número de hogares particulares con al menos 1 emigrante	43 240
Número de hogares particulares con al menos 1 emigrante y 1 inmigrante	13 046
Porcentaje de hogares particulares con al menos 1 inmigrante	16,11
Porcentaje de hogares particulares con al menos 1 emigrante	3,50
Porcentaje de hogares con al menos 1 emigrante y 1 inmigrante	1,05

Otra característica de interés de las poblaciones es la distribución en las zonas urbana y rural, de la cual se hará uso más adelante en cálculos relacionados a la muestra. Se obtuvo un 72,9% de los inmigrantes residían en la zona urbana, mientras que para los emigrantes esta cifra fue de 76,3% (cuadro 2).

Cuadro 2
Distribución de la población emigrante e inmigrante de Costa Rica en las zonas urbana y rural

Zona	Inmigrantes	Emigrantes
Urbano	72,9	76,3
Rural	27,1	23,7
Total	100	100

A pesar de que se pretendía que la encuesta tuviera cobertura nacional en función de la realización del trabajo de campo no todas las unidades primarias de muestreo fueron incluidas en los marcos muestrales, ya que en algunos distritos la cantidad de hogares con emigrantes o inmigrantes es muy baja aumentando así costos y duración del levantamiento de la información. Se incluyeron entonces, solamente aquellas ubicadas en distritos en los que la concentración de hogares con emigrantes e inmigrantes se consideró alta, sin que esto significara una pérdida excesiva en la cobertura.

En el caso de los inmigrantes se excluyeron aquellas unidades primarias de muestreo ubicadas en distritos en los que menos del 10% de los hogares son hogares inmigrantes. Este marco muestral se compuso de 111 de los 472 distritos del país, que corresponde a

una cobertura del 71,11% del total de inmigrantes que residen en Costa Rica. Algunos detalles de este marco se muestran en el cuadro 3.

Cuadro 3
Marco Muestral de hogares y personas Inmigrantes

Característica	Cifra
Número de distritos	111
Número de personas	2 257 153
Número de personas inmigrantes	271 821
Número de personas emigrantes	34 928
Número de hogares	655 015
Número de hogares con al menos 1 inmigrante	138 994
Número de hogares con al menos 1 emigrante	25 026
Porcentaje de personas inmigrantes en el marco	12,04
Porcentaje de personas emigrantes en el marco	1,5
Cobertura de personas inmigrantes del país (%)	71,11
Cobertura de personas emigrantes del país (%)	58,98
Porcentaje de hogares inmigrantes en el marco	22,11
Porcentaje de hogares emigrantes en el marco	3,82
Cobertura de hogares inmigrantes del país (%)	65,60
Cobertura de hogares emigrantes del país (%)	57,88

En el caso de la población emigrante, los porcentajes de hogares con al menos 1 emigrante en los distritos son mucho menores en comparación con la población inmigrante. Por esta razón, y nuevamente, en función del trabajo de campo, se excluyeron las unidades primarias de muestreo en aquellos distritos en los que el porcentaje de hogares con al menos 1 emigrante es menor al 3%. Esto corresponde a 193 distritos que se traduce en una cobertura de 69,01% del total de personas emigrantes costarricenses. Los detalles de este marco se muestran en el cuadro 4.

Cuadro 4
Marco Muestral de hogares y personas Emigrantes

Característica	Cifra
Número de distritos	193
Número de personas	2 081 451
Número de personas emigrantes	40 863
Número de personas inmigrantes	219 195
Número de hogares	614 518
Número de hogares con al menos 1 emigrante	29 834
Número de hogares con al menos 1 inmigrante	112 949
Porcentaje de personas emigrantes en el marco	1,96
Porcentaje de personas inmigrantes en el marco	10,53
Cobertura de personas emigrantes del país (%)	69,01
Cobertura de personas inmigrantes del país (%)	57,34
Porcentaje de hogares emigrantes en el marco	4,85
Porcentaje de hogares inmigrantes en el marco	18,38
Cobertura de hogares emigrantes del país (%)	69,00
Cobertura de hogares inmigrantes del país (%)	53,30

Se evaluó inicialmente el traslape entre los marcos muestrales y las muestras mediante el número de distritos comunes entre ambos marcos muestrales. Es decir aquellos distritos en los que la concentración de hogares inmigrantes y emigrantes es lo suficientemente alta para estar en ambos marcos muestrales. Se encontró que 63 distritos presentan esta condición (anexo 1).

Para el marco muestral de inmigración esto representa un casi un 57% de su totalidad de distritos. Debido a esto, y a que, según dicta la experiencia es más difícil captar la población emigrante; se diseñó una sola muestra con el marco muestral de emigrantes, con la cual se examina el número de esperado de hogares inmigrantes, teniendo en mente que en caso de ser necesario se integraría una muestra complementaria de inmigrantes.

Con la finalidad de mejorar los procedimientos de selección y la ejecución del trabajo de campo, se estratificó el marco muestral en 3 estratos, de acuerdo al porcentaje de hogares en los distritos con al menos 1 emigrante. Es decir, se realizó la estratificación con base en los distritos, no en las UPM.

La estratificación se realizó arbitrariamente pero fue confirmada mediante un análisis de conglomerados usando el procedimiento K-medias en el paquete STATA (anexo 2).

Los estratos construidos fueron: estrato de alta emigración, en el que se incluyeron aquellos distritos en los que el porcentaje de hogares con al menos 1 emigrante, era igual o superior a 7%, el estrato de emigración media, que contiene los distritos con porcentajes de hogares emigrantes mayores o iguales a 5% y menores que 7% y finalmente el estrato de baja emigración en el que se incluyeron los distritos en el marco con menos de 5% de hogares emigrantes. El detalle de la estratificación se muestra en el cuadro 5.

Cuadro 5
Estratificación del marco muestral de emigrantes

Estrato	Condición	Distritos	Número de		% de personas emigrantes	% de hogares emigrantes	Número de personas inmigrantes	Número de hogares con inmigrantes
			Número de personas emigrantes	hogares con emigrantes				
Alto	≥7%	26	8 211	6 065	20,09	20,33	23 593	12594
Medio	≥5% & <7%	44	12 652	9 278	30,96	31,10	59 369	31178
Bajo	≥3% & <5%	123	20 000	14 491	48,94	48,57	136 233	69177
Total		193	40 863	29 834	100,00	100,00	219 195	112949

4. Diseño de la muestra

Con los recursos disponibles se espera entrevistar cerca de 2 000 hogares con inmigrantes o emigrantes. Inicialmente se estableció el número de hogares a entrevistar para cada una de las poblaciones. La distribución del total de hogares a entrevistar entre ellas se realizó de forma arbitraria, tomando en cuenta la experiencia de encuestas previas y el hecho que la captación de emigrantes es más difícil y más costosa.

De esta forma se estableció que se esperan 800 entrevistas a hogares emigrantes y 1 200 a hogares inmigrantes. Se asumió una no respuesta de un 20% por lo que se planea la ejecución de la encuesta con 1 500 entrevistas a hogares inmigrantes y 1 000 a hogares emigrantes, dado que:

$$n_{emigrantes} = \frac{800}{0,80} = 1\ 000$$

$$n_{inmigrantes} = \frac{1\ 200}{0,80} = 1\ 500$$

Antes de realizar la afijación de estas muestras a sus estratos respectivos, se calcularon los esfuerzos para encontrar un hogar emigrante en el marco muestral, así como el número final de entrevistas esperadas para cada UPM.

El cálculo de los esfuerzos se realizó tomando en cuenta el número total de hogares en cada estrato y el número de hogares con emigrantes, de la siguiente manera:

$$esfuerzo_i = \frac{\text{total de hogares en el estrato } i}{\text{total de hogares emigrantes en el estrato } i}$$

Con esta información se calculó el número de entrevistas esperadas para cada estrato. Se sabe que las UPM construidas con base en el censo de 2011 se componen en promedio de 125 viviendas en el área urbana y 100 en el área rural (INEC, 2014), por lo que el número de entrevistas esperadas se calculó como:

$$entrevistas_i = \frac{\text{esfuerzo en el estrato } i}{\text{número de viviendas en la UPM}}$$

Los detalles de esfuerzo y número de entrevistas esperadas en las UPM se muestran en el cuadro 6.

Cuadro 6
Medida de esfuerzo y número de entrevistas esperadas por zona en el marco muestral

	Esfuerzo	Entrevistas esperadas	
		Urbano	Rural
Alto	12	11	8
Medio	17	7	6
Bajo	27	5	4

Utilizando esta información se procedió a afijar la muestra a los estratos. Se realizaron varias pruebas con diferentes afijaciones para determinar cómo ésta podría afectar el número final de UPMs necesarias, y elegir la afijación más adecuada de acuerdo a los recursos disponibles. La distribución de la muestra entre las áreas urbana y rural se realiza de acuerdo a la información presentada al inicio de este documento. Los resultados de este ejercicio se muestra en el cuadro 7.

Cuadro 7

Muestra de emigrantes: Afijación de la muestra y número de UPM necesarias en cada estrato

Afijación 1							
Estrato	% Afijación	Entrevistas	Entrevistas por UPM		UPMs necesarias		Total de UPMs
			Urbano	Rural	Urbano	Rural	
Alto	60	600	11	8	43	17	
Medio	30	300	7	6	32	12	
Bajo	10	100	5	4	17	6	127
Total	100	1000			92	35	
Afijación 2							
Estratos	% Afijación	Entrevistas	Entrevistas por UPM		UPMs necesarias		Total de UPMs
			Urbano	Rural	Urbano	Rural	
Alto	40	400	11	8	29	11	
Medio	40	400	7	6	42	16	
Bajo	20	200	5	4	32	13	143
Total	100	1000			103	40	
Afijación 3							
Estratos	% Afijación	Entrevistas	Entrevistas por UPM		UPMs necesarias		Total de UPMs
			Urbano	Rural	Urbano	Rural	
Alto	35	350	11	8	25	10	
Medio	35	350	7	6	37	14	
Bajo	30	300	5	4	48	19	153
Total	100	1000			110	43	
Afijación 4							
Estrato	% Afijación	Entrevistas	Entrevistas por UPM		UPMs necesarias		Total de UPMs
			Urbano	Rural	Urbano	Rural	
Alto	20,33	203	11	8	15	6	
Medio	31,10	311	7	6	33	13	
Bajo	48,57	486	5	4	78	30	175
Total	100	1000			126	49	

En vista de que, como ya mostró, el número de entrevistas esperadas es drásticamente menor en los estratos bajos de ambos marcos muestrales, se decide afijar la muestra intencionalmente recargada en los estratos alto y medio, es decir, seleccionar la Afijación 1.

Ahora bien, fue necesario evaluar el número de entrevistas a hogares inmigrantes que se esperan obtener mediante la muestra a emigrantes, con el fin de determinar si una muestra complementaria de hogares inmigrantes debe ser seleccionada. Se examinó lo anterior considerando nuevamente la medida de esfuerzo y el número de entrevistas esperadas para cada UPM de forma cruzada (cuadro 8).

Cuadro 8
Número de entrevistas a hogares inmigrantes por UPM esperadas en el marco de emigración

Estrato	Condición	Número de hogares con inmigrantes	Número de hogares	Esfuerzo	Entrevistas por UPM	
					Urbano	Rural
Alto	$\geq 7\%$	13 341	71 441	5	23	19
Medio	$\geq 5\% \ \& \ < 7\%$	32 800	160 577	5	26	20
Bajo	$\geq 4\%$	73 029	382 500	5	24	19

Como se muestra en el cuadro 9, solamente tomando en cuenta las 126 UPM de la muestra a emigrantes, se supera con creces la muestra deseada para inmigrantes. Esto evidencia dos aspectos importantes: el primero es que no es necesario realizar una muestra de inmigrantes, y el segundo es que es necesario seleccionar los hogares inmigrantes con el fin de llegar al número deseado de entrevistas.

Cuadro 9
Muestra de emigrantes: Número de hogares inmigrantes esperados en la muestra

Estrato	Condición	Entrevistas por UPM		UPM		Número de entrevistas		Total
		Urbano	Rural	Urbano	Rural	Urbano	Rural	
Alto	$\geq 7\%$	23	19	43	17	1 007	313	
Medio	$\geq 5\% \ \& \ < 7\%$	26	20	32	12	809	251	2 884
Bajo	$\geq 3\% \ \& \ < 5\%$	24	19	16	6	385	119	

Con la finalidad reducir la cantidad de entrevistas que se llevarían a cabo en hogares inmigrantes se realizará un muestreo sistemático para muestrear solamente las UPM necesarias para obtener el tamaño de muestra inicialmente planteado. Es decir el módulo de la encuesta planteado para captar la información de los hogares con inmigrantes no será aplicado en las 127 UPM inicialmente seleccionadas, sino en una submuestra de ellas.

Para determinar la estrategia de muestreo adecuada en primer lugar se evaluó la concentración de hogares inmigrantes en los distritos incluidos en cada uno de los estratos. Se encontró que la el porcentaje promedio de hogares inmigrantes entre los estratos no eran muy distintos entre sí (cuadro 10), por lo que no se considera que se requiera una afijación que recargue la muestra en algún estrato para la selección de la muestra.

Cuadro 10
Porcentaje promedio de hogares inmigrantes en los distritos por los estratos de emigración

Estrato	Condición	Porcentaje promedio de hogares inmigrantes
Alto	≥7%	16,98
Medio	≥5% & <7%	19,42
Bajo	≥3% & <5%	18,21

Sin embargo, considerando las características anteriormente presentadas sobre la distribución de las UPM a seleccionar en los estratos de emigración, se asignan menos entrevistas al estrato bajo con el fin de que las UPM extraídas para la muestra de emigración sean suficientes para realizar la selección de la muestra de inmigración en él. La distribución de las UPM a seleccionar para la muestra de inmigración se muestra en el cuadro 11.

Cuadro 11
Número de UPMs necesarias para la muestra de inmigrantes a extraer de la muestra de emigrantes

Estrato	Condición	Entrevistas por UPM			UPMs necesarias	
		Entrevistas	Urbano	Rural	Urbano	Rural
Alto	>=7%	600	22	18	21	8
Medio	>=5% & <7%	600	24	19	19	7
Bajo	>=4%	300	23	18	10	4

En vista de que se cuenta con dos muestras distintas, las fórmulas de selección para ambas son distintas. En el caso de la muestra de emigración, se tiene que:

$$f_{emig} = \frac{a}{A} \times \frac{B}{B}$$

Donde,

- A corresponde al número total de UPMs en el marco
- a corresponde al número de UPMs seleccionadas en la muestra
- B corresponde al tamaño de la UPM

El procedimiento de selección de las UPM dentro de los estratos para la muestra de emigrantes se llevará a cabo utilizando muestreo sistemático dentro de cada estrato, por lo que el espaciamiento k_{emig} está definido como:

$$k_{emig} = \frac{\text{Número de UPM en el estrato}}{\text{Número de UPM a seleccionar en el estrato}}$$

La muestra de inmigrantes por otro lado se obtiene de la muestra inicialmente obtenida de para el estudio de la emigración. En este caso, se seleccionan el número de UPMs necesarias en cada estrato para obtener el tamaño de muestra deseado. El número de UPMs a seleccionar para la muestra de inmigración por cada UPM en la muestra de emigración (m) se define como:

$$m = \frac{\text{Número de UPM seleccionadas para emigración}}{\text{Número de UPM a seleccionar para inmigración}}$$

Los resultados del cálculo de m para cada estrato muestral se presentan el cuadro 12. Nótese que todas las cifras son cercanas a 0,5, es decir, para cada estrato se seleccionarán para la muestra de hogares inmigrantes aproximadamente la mitad de UPMs seleccionadas originalmente para la muestra de hogares emigrantes.

Cuadro 12
Espaciamientos para la selección de las UPM en la muestra de inmigración

Estrato	Condición	Espaciamiento	
		Urbano	Rural
Alto	$\geq 7\%$	0,49	0,47
Medio	$\geq 5\% \ \& \ < 7\%$	0,59	0,58
Bajo	$\geq 4\%$	0,62	0,66

La fórmula de selección para la muestra de hogares inmigrantes, queda definida como:

$$f_{emig} = m \frac{a}{A} \times \frac{B}{B}$$

Donde,

- A corresponde al número total de UPMs en el marco
- a corresponde al número de UPMs seleccionadas en la muestra
- B corresponde al tamaño de la UPM
- m corresponde al número de UPMs a seleccionar para la muestra de hogares inmigrantes por cada UPM seleccionada para la muestra de hogares emigrantes.

Cálculo de errores muestrales e intervalos de confianza

El objetivo de una encuesta casi que sin excepción realizar inferencias sobre la población de estudio. Debido a que las muestras diseñadas son probabilísticas, son también medibles, de forma que es posible calcular medidas de variabilidad con la información recolectada, estas medidas son los errores muestrales y permiten a su vez la estimación de intervalos de confianza para las estimaciones.

Ambas muestras son a su vez muestras complejas, por lo que el cálculo de las variancias difiere de aquél que se efectuaría si se realizara un muestreo irrestricto aleatorio (mia). En este caso suponiendo una muestra apropiadamente ponderada el cálculo de la varianza para una proporción queda dado por:

$$var(p) = \sum_{h=1}^H W_h^2 (1 - f_h) \frac{p_h q_h}{n_h - 1}$$

Dónde

- p_h corresponde a la proporción de interés en cada estrato de selección
- q_h corresponde al complemento de la proporción de interés en cada estrato de selección
- n_h corresponde al tamaño de muestra en cada estrato de selección
- f_h corresponde a la fracción de muestreo en cada estrato
- W_h corresponde al peso del estrato

Y el intervalo de confianza queda definido por:

$$p \pm z_{\alpha/2} \sqrt{1 - f} \sqrt{var(p)} + \frac{1}{2n}$$

Donde $z_{\alpha/2}$ representa el valor de una distribución de probabilidad normal estándar para un nivel de confianza definido, f representa la fracción de muestreo general y n el tamaño de muestra total.

En el caso de una variable métrica, se la fórmula de la varianza, nuevamente asumiendo una muestra apropiadamente ponderada es:

$$var(\bar{y}) = \frac{(1 - f)}{n^2} \sum_{h=1}^H n_h s_h^2$$

Dónde

- \bar{y} representa la estimación obtenida para el promedio de la variable de interés
- n_h corresponde al tamaño de muestra en cada estrato de selección
- s_h^2 corresponde la varianza de la variable en cada estrato de selección
- n corresponde al tamaño de muestra total
- f representa la fracción de muestreo general

Análogamente el intervalo de confianza se define como:

$$\bar{y} \pm z_{\alpha/2} \sqrt{1 - f} \sqrt{var(\bar{y})}$$

Bajo las mismas definiciones ya mencionadas.

Limitaciones de la muestra

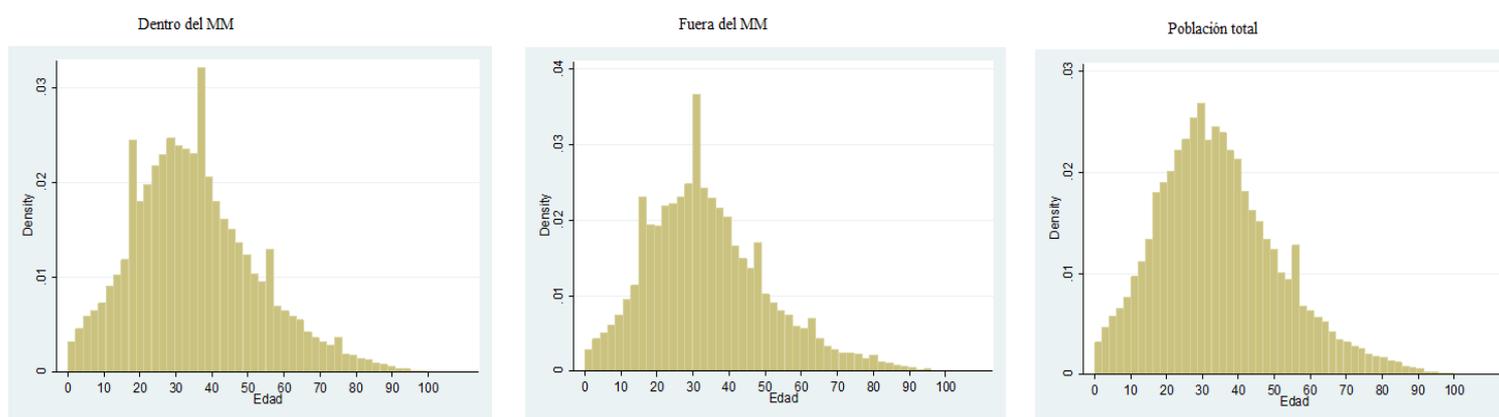
El uso del censo de población y vivienda de 2011 como base para la construcción del marco muestral presenta algunos inconvenientes. Primero, para la población emigrante solamente se cuenta con el número de personas que han dejado el hogar; impidiendo conocer características propias de la población, y segundo, para la población inmigrante se estima que la omisión censal fue cercana al 22% (CCP, 2013); no obstante, no se considera que alguna otra fuente brinde información adecuada sobre esta temática.

La limitación principal del diseño propuesto se encuentra en la cobertura de las poblaciones objetivo. La cobertura del marco muestral de emigración, que es el utilizado para la selección de la muestra de hogares inmigrantes y emigrantes cubre un 70% de las personas emigrantes y cerca de un 60% de las personas inmigrantes.

Para medir el efecto que esto podría tener en la muestra final, se comparan la composición de algunas distribuciones de las poblaciones objetivo con las del marco muestral de emigración en variables de composición demográfica para el caso de las personas inmigrantes. Por otro lado, debido a la limitación de la información, para la población emigrante se compara la distribución porcentual de algunas variables de interés; pero, a nivel del hogar. Los resultados se muestran en los gráficos 1 y 2 y los cuadros 13 y 14.

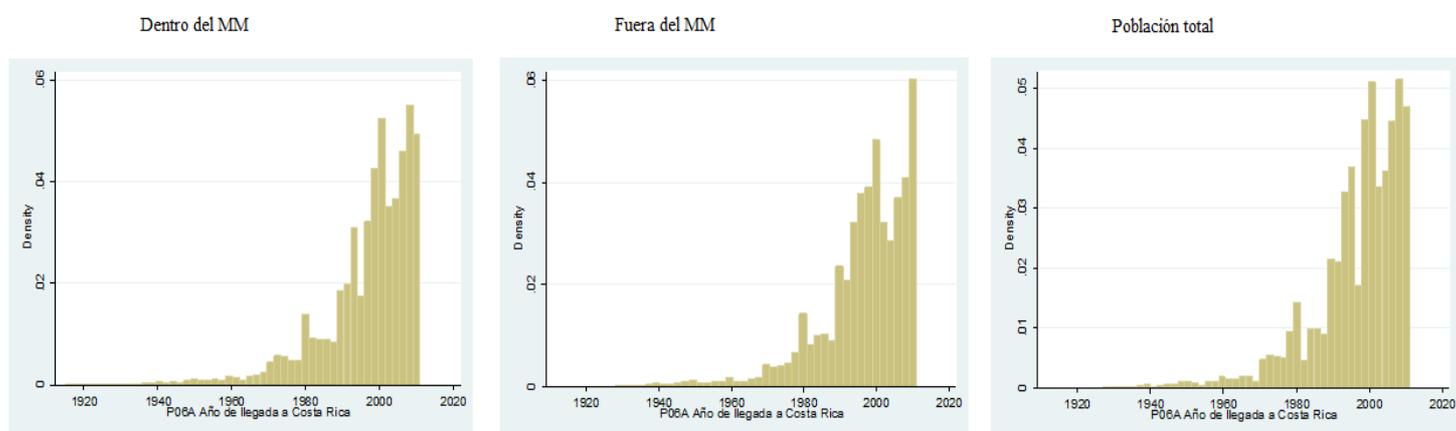
En cuanto a la edad, se observa concentración de personas cerca de los 20 y 40 años; sin embargo, el resto de la distribución se asemeja a la de la población total.

Gráfico 1
Comparación de la distribución de la edad en para las personas inmigrantes: dentro del MM, fuera del MM y la población total



Otra de las variables de interés es el año de llegada a Costa Rica de las personas inmigrantes. En este caso en particular, como se muestra en el gráfico 2, no se observan grandes desviaciones de la población total tanto para las personas dentro del MM como aquellas fuera de él.

Gráfico 2
Comparación de la distribución del año de llegada a Costa Rica para las personas inmigrantes: dentro del MM, fuera del MM y la población total



También fueron examinadas las distribuciones porcentuales del nivel educativo, la nacionalidad, la zona de residencia, el tamaño promedio del hogar, la categoría ocupacional, y el estado conyugal. Lo encontrado sugiere que en el marco muestral se incluyó una proporción menor de personas con nivel educativo bajo (42,23%) en comparación con la población total (50,22%), mientras que el porcentaje de personas con formación universitaria (19,31%) es mayor que el encontrado en la población (14,63%).

Se encontraron algunas diferencias en la composición de las nacionalidades, ya que el porcentaje de personas nicaragüenses en el marco muestral (66,52%) es menor que el de la población total (74,67%). Otra diferencia se presentó en la zona de residencia, con una concentración mayor de personas residentes en la zona urbana en el marco muestral; sin embargo, en este caso particular cabe rescatar que la muestra fue afijada de acuerdo con la distribución urbana y rural de la población total de personas inmigrantes.

Cuadro 13
Comparación de la población inmigrante en el Marco Muestral, fuera de él y la población total

Característica	Dentro del MM	Fuera del MM	Población total
<i><u>Nivel educativo</u></i>			
Primaria o menos	42,23	60,96	50,22
Secundaria académica técnica o parauniversitaria	34,14	27,39	31,26
Universitaria	4,32	3,30	3,89
	19,31	8,35	14,63
<i><u>Nacionalidad</u></i>			
China	1,01	0,64	0,85
Colombia	5,96	1,99	4,26
Cuba	1,33	0,57	1,01
El Salvador	2,75	2,01	2,43
Estados Unidos	5,27	2,63	4,15
Honduras	1,09	0,79	0,96
México	1,10	0,34	0,78
Nicaragua	66,52	85,62	74,67
Otros	8,31	3,11	6,09
Panamá	3,91	1,55	2,90
Perú	1,23	0,43	0,88
Venezuela	1,53	0,31	1,01
<i><u>Zona de residencia</u></i>			
Urbano	84,26	57,51	72,85
Rural	15,74	42,49	27,15
<i><u>Tamaño del hogar</u></i>			
Número promedio de personas en el hogar	4,33	4,61	4,45
<i><u>Sexo</u></i>			
Hombre	46,99	49,37	48,01
Mujer	53,01	50,63	51,99
<i><u>Categoría ocupacional</u></i>			
Patrono(a)	8,61	4,83	7,10
Trabajador(a) por cuenta propia	19,01	19,42	19,17
Empleado(a) de empresa privada	55,70	64,26	59,12
Empleado(a) del sector público	3,81	3,18	3,55
Empleado(a) de casas particulares	12,08	7,28	10,16
Ayudante sin recibir pago	0,79	1,04	0,89
<i><u>Estado conyugal</u></i>			
Unión libre o juntado(a)	29,99	38,69	33,71
Casado(a)	31,61	26,70	29,51
Separado(a)	3,78	3,70	3,74
Divorciado(a)	2,87	1,40	2,24
Viudo(a)	2,66	2,29	2,50
Soltero(a)	29,1	27,23	28,30

Como se especificó, el análisis para el caso de la población emigrante se hizo al nivel agregado del hogar (cuadro 14). En este caso no se encontraron diferencias claras entre los hogares incluidos en el marco muestral y la población total de hogares con personas emigrantes. Variables como la zona de residencia, el sexo de la jefatura del hogar, el promedio de años de escolaridad de la jefatura, su edad promedio y condición de actividad mostraron distribuciones similares.

Cuadro 14
Comparación de la población emigrante en el Marco Muestral y la población total

Característica	Dentro del MM	Población total
Zona		
Urbano	80,94	76,29
Rural	19,06	23,71
Sexo del jefe(a)		
Hombre	59,82	61,33
Mujer	40,18	38,67
Promedio de años de escolaridad del jefe(a) del hogar		
	9,63	9,09
Condición de actividad del jefe(a) del hogar		
Ocupados	58,75	59,71
Desocupado buscó trabajo y había trabajado	1,01	1,08
Desocupado buscó trabajo primera vez	0,07	0,06
Inactivo pensionado(a) o jubilado(a)	15,92	14,92
Inactivo vive de rentas o alquileres	0,89	0,83
Inactivo sólo estudia	0,98	0,90
Inactivo se dedica a oficios domésticos	14,78	14,83
Inactivo otra situación	7,61	7,68
Promedio de edad del jefe(a) del hogar		
	53,60	52,78

4. Simulación de la selección de la muestra

Con el marco muestral que se posee se evalúa el diseño de la muestra y las características de la misma. El marco muestral con el que se cuenta no permite obtener información a nivel de UPM por lo que no se pudo realizar una selección directa de las mismas sino que fue necesario reconstruirlas.

Sabiendo que las UPM contienen en promedio 100 viviendas en el área urbana y 125 en el área rural, se ordenaron las UGM de acuerdo a el distrito y la zona y se agruparon en conjuntos de 3 en la zona urbana y 9 en la zona rural para obtener un número de viviendas

que se acercara a estos promedios. El promedio de viviendas por UPM por zona de residencia en el marco muestral se presenta en el cuadro 15.

Cuadro 15
Promedio de viviendas por UPM reconstruida según zona de residencia

Zona de residencia	Promedio
Urbano	99,04
Rural	115,66

Utilizando el diseño de muestra aquí planteado se seleccionaron 127 de las UPM reconstruidas. Con esta muestra se obtuvieron el número esperado de hogares con emigrantes e inmigrantes (cuadro 16). Se observa que los esperados para ambos tipos de hogares son levemente mayores de los obtenidos mediante la muestra; sin embargo, se considera que esto se debe a la reconstrucción de las UPM y no al diseño, dado que las UPM reconstruidas son ligeramente más pequeñas que las planteadas en el diseño.

Cuadro 16
Número de hogares con migrantes en las 127 UPM seleccionadas

Tipo de hogar	Número de hogares
Con emigrantes	890
Con inmigrantes	2 170

Nótese que en el cuadro anterior no se filtraron las UPM para la muestra de inmigrantes. Esto se realizó posterior a la selección, ordenando las UPM por estrato de emigración y zona, y se seleccionaron sistemáticamente las UPM a conservar de acuerdo con el número de UPM necesarias por estrato mostrado en el apartado anterior. Después de realizar este filtro se obtuvieron 1 300 hogares con inmigrantes y 890 con emigrantes (cuadro 17).

Cuadro 17
Número de hogares en la muestra final de personas emigrantes e inmigrantes

Tipo de hogar	Número de hogares
Con emigrantes	890
Con inmigrantes	2 170

5. El caso inverso: planteando la muestra desde el marco muestral de inmigración

Todos los procedimientos que se han mostrado hasta este punto provienen del marco muestral de emigración y no se ha mostrado los efectos que tendría la muestra si se planteara el diseño desde el marco muestral de inmigración. Se procede en esta sección a realizar justamente esto.

Lo primero es el planteamiento de la estratificación, de forma análoga al marco muestral de emigrantes se construyeron 3 estratos de distritos arbitrariamente, también verificados mediante análisis de conglomerados utilizando el procedimiento de K-medias (anexo 3). En el estrato de alta inmigración se incluyeron todos aquellos distritos con al menos 30% de hogares con 1 inmigrante o más, en el de inmigración media, todos aquellos con al menos 20% y menos de 30% hogares inmigrantes y finalmente en el de baja inmigración se incluyeron todos los distritos cuyo porcentaje de hogares inmigrantes fue mayor al 5% pero no superó el 20% (cuadro 18).

Cuadro 18
Estratificación del marco muestral de inmigrantes

Estrato	Condición	Distritos	% de distritos	Número de personas inmigrantes	Número de hogares con inmigrantes	% de personas inmigrantes	% de hogares inmigrantes
Alto	≥30%	28	25,23	83 008	40 378	30,54	27,86
Medio	≥20% & <30%	39	35,14	108 087	56 721	39,76	39,14
Bajo	≥5% & <20%	44	39,63	80 726	47 835	29,70	33,00
Total		111	100,00	271 821	144 934	100,00	100,00

El siguiente paso fue calcular la medida de esfuerzo para encontrar un hogar con inmigrantes y el número de entrevistas esperadas por UPM, bajo el conocimiento de que el tamaño promedio de las UPM en la zona urbana es de 100 viviendas y en la zona rural es de 124 viviendas. En el cuadro 19 se aprecia que en comparación al número de entrevistas de hogares inmigrantes que se obtienen bajo el marco muestral de emigrantes, este marco muestral produce un número mayor de entrevistas. De hecho y como se mostró anteriormente mediante la concentración de inmigrantes en cada estrato del marco muestral de emigración, el diseño de muestra ya presentado puede considerarse como un equivalente al muestreo del estrato bajo del marco muestral de inmigración.

Cuadro 19
Medida de esfuerzo y número de entrevistas esperadas por zona en el marco muestral de inmigración

Estrato	Condición	Esfuerzo	Entrevistas esperadas	
			Urbano	Rural
Alto	≥30%	2	51	41
Medio	≥20% & <30%	4	31	24
Bajo	≥5% & <20%	7	18	15

Análogamente a lo presentado en apartados anteriores, utilizando esta información se procedió a afijar la muestra a los estratos. Se realizaron nuevamente varias pruebas con

diferentes afijaciones para determinar cómo ésta podría afectar el número final de UPMs necesarias, correspondiendo la última afijación probada a una afijación proporcional. En el cuadro 20 se evidencia que el número total de UPMs necesarias para una muestra efectiva de hogares inmigración es mucho menor en comparación de la muestra de hogares emigrantes.

Cuadro 20
Muestra de inmigrantes: Afijación de la muestra y número de UPM necesarias en cada estrato

Afijación 1							
Estrato	% Afijación	Entrevistas	Entrevistas por UPM		UPMs necesarias		Total de UPMs
			Urbano	Rural	Urbano	Rural	
Alto	60	900	51	41	13	6	
Medio	30	450	31	24	11	5	
Bajo	10	150	18	15	6	3	
Total	100	1500			30	14	43
Afijación 2							
Estrato	% Afijación	Entrevistas	Entrevistas por UPM		UPMs necesarias		Total de UPMs
			Urbano	Rural	Urbano	Rural	
Alto	40	600	51	41	9	4	
Medio	40	600	31	24	14	7	
Bajo	20	300	18	15	12	6	
Total	100	1500			35	16	51
Afijación 3							
Estrato	% Afijación	Entrevistas	Entrevistas por UPM		UPMs necesarias		Total de UPMs
			Urbano	Rural	Urbano	Rural	
Alto	35	420	51	41	6	3	
Medio	35	420	31	24	10	5	
Bajo	30	360	18	15	14	7	
Total	100	1200			30	14	44
Afijación 4							
Estrato	% Afijación	Entrevistas	Entrevistas por UPM		UPMs necesarias		Total de UPMs
			Urbano	Rural	Urbano	Rural	
Alto	30,79	369	51	41	5	2	
Medio	40,09	481	31	24	11	5	
Bajo	29,12	349	18	15	14	6	
Total	100	1200			31	14	45

El próximo paso fue determinar si con este número de UPM es posible obtener las entrevistas para la muestra a hogares emigrantes. Esto se muestra en el cuadro 21. El número de entrevistas esperadas a hogares emigrantes es muy bajo, menos de 250 entrevistas en todos los casos de afijación. Esto implicaría planear una muestra de emigrantes aparte, de la que como ya se mostró es posible obtener un alto número de entrevistas a hogares inmigrantes, por lo que este enfoque se vuelve poco práctico.

Cuadro 21
Muestra de inmigrantes: Número de hogares emigrantes esperados en la muestra

Afijación 1								
Estrato	Condición	Entrevistas por UPM		UPMs		Total de entrevistas		Total
		Urbano	Rural	Urbano	Rural	Urbano	Rural	
Alto	$\geq 30\%$	6	5	13	6	74	27	
Medio-Alto	$\geq 20\% \ \& \ < 30\%$	5	4	11	5	50	19	
Medio	$\geq 10\% \ \& \ < 20\%$	4	4	6	3	26	10	206
Afijación 2								
Estrato	Condición	Entrevistas por UPM		UPMs		Total de entrevistas		Total
		Urbano	Rural	Urbano	Rural	Urbano	Rural	
Alto	$\geq 30\%$	6	5	9	4	49	18	
Medio-Alto	$\geq 20\% \ \& \ < 30\%$	5	4	14	7	67	25	
Medio	$\geq 10\% \ \& \ < 20\%$	4	4	12	6	52	19	231
Afijación 3								
Estrato	Condición	Entrevistas por UPM		UPMs		Total de entrevistas		Total
		Urbano	Rural	Urbano	Rural	Urbano	Rural	
Alto	$\geq 30\%$	6	5	6	3	34	13	
Medio-Alto	$\geq 20\% \ \& \ < 30\%$	5	4	10	5	47	17	
Medio	$\geq 10\% \ \& \ < 20\%$	4	4	14	7	63	23	197
Afijación 4								
Estrato	Condición	Entrevistas por UPM		UPMs		Total de entrevistas		Total
		Urbano	Rural	Urbano	Rural	Urbano	Rural	
Alto	$\geq 30\%$	6	5	5	2	30	11	
Medio-Alto	$\geq 20\% \ \& \ < 30\%$	5	4	11	5	54	20	
Medio	$\geq 10\% \ \& \ < 20\%$	4	4	14	6	61	23	199

6. Anexos

Anexo 1 Distritos comunes entre los marcos muestrales (anexo)

Número de Distrito	
10102	11502
10103	11504
10104	11801
10105	11802
10106	11804
10107	11901
10108	20101
10109	20102
10110	20106
10111	20109
10201	20110
10202	20306
10203	21007
10301	30302
10302	30303
10311	40101
10701	40102
10801	40103
10803	40104
10804	40703
10805	40901
10901	41002
10903	50309
10904	50503
10905	60111
11004	61003
11005	61101
11301	70101
11302	70402
11305	70403
11401	
11403	
11501	

Anexo 2

Estratificación del marco muestral de emigrantes utilizando análisis de conglomerados

Estrato	Condición	Distritos	% de distritos	Número de personas emigrantes	Número de hogares con emigrantes	% de personas emigrantes	% de hogares emigrantes
Alto	$\geq 7,82\%$ & $\leq 12,06\%$	23	11,92	6 781	5 019	16,59	16,82
Medio	$\geq 4,69\%$ & $\leq 7,38\%$	58	30,05	16 485	12 072	40,34	40,46
Bajo	$\geq 3,01\%$ & $\leq 4,66\%$	112	58,03	17 597	12 743	43,07	42,72
Total		193	100,00	40 863	29 834	100,00	100,00

Anexo 3

Estratificación del marco muestral de inmigrantes utilizando análisis de conglomerados

Estrato	Condición	Distritos	% de distritos	Número de personas inmigrantes	Número de hogares con inmigrantes	% de personas inmigrantes	% de hogares inmigrantes
Alto	$\geq 35,29\%$ & $\leq 54,57\%$	21	18,92	66 301	32 098	24,39	22,15
Medio	$\geq 20,81\%$ & $\leq 33,16\%$	44	39,64	119 169	61 952	43,84	42,74
Bajo	$\geq 6,26\%$ & $\leq 20,24\%$	46	41,44	86 351	50 884	31,77	35,11
Total		111	100,00	271 821	144 934	100,00	100,00

6. Referencias

CCP. (2013). " *Evaluación demográfica del X Censo Nacional de Población de Costa Rica 2011 y de otras fuentes de información*". San José.

INEC. (2014). " *Nueva muestra a partir del Marco Muestral de Viviendas 2011 y Proyecciones de Población 2013*". San José.