



Anexo II

METODOLOGÍA DE ENCUESTA Y MUESTREO



Por universo muestral se entiende la “totalidad de individuos o elementos en los cuales puede presentarse determinada característica susceptible a ser estudiada por medio de una encuesta”. En la versión original del estudio, el universo muestral estaba conformado por el conjunto de la población que reside en las Zonas de Medios de Vida (ZMV) donde opera CRS y entidades socias, esto es, GT05, GT06, GT07, GT10, GT11. Este universo muestral, en el que se asientan miles de comunidades, implicaba que el equipo de investigación elaborara con información propia, del censo de población (INE, 2002 y 2018) e insumos proporcionados por CRS y entidades socias, una Base de Datos (BBDD) en la que incluir el nombre de todas las comunidades y núcleos urbanos censados, su población estimada, la presencia de proyectos de CRS en algunas de estas comunidades, y el número de participantes en tales proyectos. Sobre esta BBDD se aplicaría un diseño muestral que contemplara una estratificación en función de las ZMV, para dar una cobertura proporcional de las comunidades donde CRS realiza algún proyecto. La estratificación garantizaba así también que los participantes a los proyectos de CRS estuviesen representados en la muestra con un tamaño suficiente y proporcional a su peso en la población. Aplicando un software de muestreo se seleccionaron al azar 20 clúster o núcleos habitados, en los que se debían realizar un total de, al menos, 623 encuestas.

Todo este ejercicio quedó sin vigencia debido a la emergencia sanitaria provocada por el COVID 19, ante la imposibilidad de realizar el trabajo de campo de manera tradicional con los equipos de encuestadores desplazándose a comunidades y hogares. En este contexto se optó por una encuesta telefónica, la cual exigía;

- i. Reducir el tamaño del cuestionario de encuesta para facilitar su aplicación vía telefónica, al tiempo que se preservan las variables más destacadas que permitan conocer mejor el proceso migratorio y de arraigo.
- ii. Establecer un nuevo protocolo de realización de la encuesta vía telefónica por medio, incluyendo actividades de comunicación y la disponibilidad de incentivos para facilitar el proceso.
- iii. Plantear un nuevo proceso de muestreo estadístico para la selección de la muestra, considerando las nuevas limitaciones respecto a la disponibilidad de contactos telefónicos.

TIPO DE MUESTREO ENCUESTA POBLACIÓN RESIDENTE EN ZMVS EN LAS QUE OPERAN PROYECTOS DE CRS

La delimitación del tamaño muestral viene definida en función de la población objetivo, participante del proceso de encuesta. Para realizar la encuesta entre habitantes de las ZMV donde opera CRS y entidades socias, se ha propuesto una estratificación en función de las ZMV, y que además da una cobertura proporcional de las comunidades donde CRS realiza algún proyecto. La estratificación garantiza también que los participantes a los proyectos de CRS estén representados en la muestra con un tamaño suficiente y proporcional a su peso en la población. Así que en los análisis de los datos será posible añadir también una variable relativa a la participación o no participación en los proyectos.

El muestreo estadístico utilizado es un **muestreo bietápico sin reemplazo, con probabilidades desiguales en ambas etapas**, para tener en cuenta el tamaño de las ZMV y de las CRs, de modo que se vean representadas proporcionalmente:

- a) En la primera etapa de este muestreo se selecciona un número de comunidades CRS dentro de las ZMV, por lo tanto se emplea un **muestreo estratificado por conglomerados**, donde los estratos son las ZMV y los conglomerados las comunidades CRS. Los estratos se caracterizan por ser heterogéneos entre sí y estar compuestos por unidades homogéneas, mientras que los conglomerados se caracterizan por estar compuestos de unidades heterogéneas y ser homogéneos entre sí. Esta diferencia es la que ocasiona que se tomen unidades de todos los estratos, mientras que no es preciso tomar unidades de todos los conglomerados.

La estratificación tiene varias ventajas. Primero, garantiza una representación adecuada de cada estrato según el peso que este estrato tiene en la población. Segundo, la estratificación garantiza una mayor eficiencia a paridad de tamaño de la muestra, sobre todo cuando la variable de estratificación está relacionada con la variable de interés como en nuestro caso. Eso quiere decir que se podría reducir el tamaño de la muestra necesario a garantizar un determinado nivel de significación y poder estadístico.

- b) En la segunda etapa se realiza la selección de personas dentro de las comunidades CRS (conglomerados) a las que se les va a pasar la encuesta, este muestreo se lleva a cabo mediante un muestreo por conveniencia, ya que los datos de contacto de esas personas son cedidos por organizaciones externas y no disponemos de el listado completo de contactos, pero la cesión de datos se realiza de forma completamente aleatoria, por lo que este tipo de muestreo funcionará igual a un muestreo aleatorio simple.

MUESTREO POR CONVENIENCIA.

El muestreo por conveniencia es un muestreo no probabilístico que se usa cuando no tenemos acceso a una lista completa de los individuos que forman la población (marco muestral) (Lavrakas, 2008). El muestreo por conveniencia es una técnica comúnmente usada consistente en seleccionar una muestra de una población por el hecho de que sea accesible, es decir, los individuos empleados en la investigación se seleccionan porque están fácilmente disponibles y porque sabemos que pertenecen a la población de interés, no porque hayan sido seleccionados mediante un criterio estadístico. Esta conveniencia, se suele traducir en una gran facilidad operativa y en bajos costes de muestreo (Sedgwick, 2013).

Lo anterior no significa que los resultados obtenidos mediante esta técnica sean irrelevantes, ya que si existen razones para pensar que la selección por conveniencia no va a introducir sesgos respecto al total de población, los resultados que obtengamos pueden ser una buena imagen del universo estudiado. El problema es que no sabremos exactamente cómo de buena imagen es, ya que no se pueden usar herramientas como el margen de error y el intervalo de confianza para medir la precisión de los resultados, por lo que hay que confiar en el criterio de selección hecha por el investigador. Algunos criterios de calidad en este diseño se abordan en (Farrokhi y Mahmoudi-Hamidabad, 2012).

Además, si el investigador considera que ese muestreo hecho por conveniencia se corresponde con algún tipo de muestreo estadístico probabilístico, se puede calcular el margen de error o el tamaño de muestra requerido para un error máximo permitido, aunque siempre hay que indicar las condiciones bajo las que se ha obtenido la muestra.

Este tipo de muestreo realmente es uno de los más utilizados, sobre todo en estudios clínicos u otros realizados con voluntarios, donde la población objeto de estudio o bien no se conoce o si se conoce, no es posible acceder a todas las personas que la componen por lo que se trabaja con voluntarios, suponiendo que esos voluntarios reúnen las mismas características que otros que no se hayan ofrecido (caso de enfermos, drogadictos, migrantes, etc.), por lo que el hecho de trabajar con los voluntarios es de suponer no introducirá un sesgo adicional en el resultado final de la investigación. Algunos ejemplos de muestreo por convención pueden encontrarse en (Murray y Hannan, 1990), (Bujang et al., 2012), (Suresh y Chandrashekara, 2012), (Brodaty et al., 2014), (Chandler y Shapiro, 2016).

1ª ETAPA: SELECCIÓN Y NÚMERO DE CONGLOMERADOS DE LA ENCUESTA DE ZMVS DONDE OPERA CRS

En Guatemala, existen 13 zonas de medios de vida, de acuerdo a la propuesta metodológica y análisis efectuado por FEWSNET en 2016, con apoyo de USAID. Según bases de datos proporcionadas por CRS de la ubicación de sus proyectos, éstos se concentran en 7 zonas de medios de vida. Sin embargo, a efectos del muestreo previsto, se excluyen dos ZMV en las que solo se implementa el proyecto Jóvenes Constructores; ZMV GT09, que corresponde a la ciudad capital y su área de influencia y GT12, donde además solo se trabaja con 5 participantes.

De acuerdo a los ZMV priorizados, se dispondrá de 5 estratos como máximo. A cada estrato se asigna una cuota de la muestra total proporcional a su dimensión (% de individuos u hogares en el estrato; “repartición proporcional”). Dentro de cada estrato, el muestreo se realiza en dos etapas. En la primera etapa se elige de forma aleatoria una muestra de comunidades o conglomerados.

Para saber cuántos conglomerados elegir de forma aleatoria, haremos uso de la fórmula para el cálculo de muestra del muestreo aleatorio simple para la estimación de la proporción, conocido el tamaño poblacional. Esta fórmula tiene la siguiente expresión:

$$n = \frac{N * Z_{\alpha}^2 * p * q}{e^2(N - 1) + Z_{\alpha}^2 * p * q}$$

donde:

N = 549 comunidades (total poblacional)

Z_{α} = 1.645 (percentil 95 de la distribución Normal ($\alpha = 0.95$))

P = 0.25 (proporción muestral de cualquier parámetro objeto de estudio cuya proporción se desconoce, por lo que se toma el valor 0.25 por ser un valor intermedio, ni el más favorable ni el más desfavorable que sería 0.5)

q = 1 – p = 0.5

e = 0.13 (error máximo admisible)

Tomando estos valores para los parámetros, el tamaño de conglomerados que hay que seleccionar es $n = 28.53$, que redondeamos a 29 conglomerados.

Todas las comunidades se tienen que repartir en los 5 estratos formados por: GT05, GT06, GT07, GT10 y GT11. La Tabla 2 representa la distribución del número de comunidades por ZMV, así como la población de las ZMV y los participantes en proyectos CRS. La columna 3 indica el peso de cada estrato en función del número de comunidades que lo componen. Para seleccionar la muestra de comunidades, multiplicaremos el número total de comunidades a seleccionar (29) por el peso de cada estrato y redondeamos al entero más cercano. De esta forma sabremos cuántas comunidades debemos seleccionar en cada uno de los estratos (columna 4). Notemos que a causa del redondeo el número de comunidades ha aumentado a 30.

DISTRIBUCIÓN DE LAS COMUNIDADES POR ZMV

ZMV	Nº COMUNIDADES CRS TOTAL	PESO	MUESTRA DE COMUNIDADES CRS
GT05	10	0,0182	1
GT06	436	0,7942	23
GT07	17	0,0310	1
GT10	38	0,0692	2
GT11	48	0,0874	3
Total	549	521626	30

2ª ETAPA: SELECCIÓN Y NÚMERO DE PARTICIPANTES FINALES DE LA ENCUESTA DE ZMVS DONDE OPERA CRS

Para el cálculo de la dimensión de la muestra se considera que el análisis principal que se utilizará consistirá en estimar un modelo de regresión logística. El modelo de regresión logística permite analizar cómo varios factores (variables independientes) están relacionados a una variable dependiente binaria. El modelo permite analizar varios factores simultáneamente y se podrá utilizar para analizar si la intención migratoria (variable binaria dependiente de interés) está en relación a varios factores (por ejemplo: zona de medio de vida, género, edad, educación, etc.).

Para el cálculo del tamaño de la muestra final, utilizamos el método propuesto por Hsieh (1989) que ha demostrado tener buenas propiedades estadísticas. Empezamos con considerar una variable única independiente X en un modelo de regresión logística. Vamos a considerar el tamaño de la muestra mínimo para garantizar un determinado poder estadístico y nivel de significación para detectar un determinado efecto de la variable independiente X sobre la intención migratoria. Medimos este efecto a través del odds ratio (OR), $OR = \frac{p_A/(1-p_A)}{p_B/(1-p_B)}$, donde pA y pB representan las probabilidades de intención migratoria para dos individuos A y B (que difieren por niveles diferentes de la variable X).

El tamaño de la muestra se obtiene con la siguiente fórmula:

$$n_U = [Z_{\alpha/2} + \exp(-\log(OR)^2 / 4) Z_{\beta}]^2 (1 + 2p\delta) / (p \log(OR)^2)$$

donde:

- $\log(OR)$ es el logaritmo del OR correspondiente a un aumento de una desviación estándar en la variable X;
- $Z_{\alpha/2}$ = valor de Z (distribución normal estándar) que corresponde a un determinado nivel de significación (α);
- Z_{β} = valor de Z (distribución normal estándar) que corresponde a un determinado poder estadístico ($1-\beta$);
- p es la proporción de individuos que tienen intención a migrar en la población;
- δ es un factor de corrección que depende de $\log(OR)$ (véase Hsieh, 1989).

Los valores fundamentales para el cálculo del tamaño de la muestra se han fijado de la siguiente forma.

- El valor p (proporción de individuos que tienen una intención de migrar en la población se fija en 0,25 (25%) que es muy parecido al valor estimado por la encuesta Gallup WorldPoll en el año 2016 (26%).
- El OR se fija a un valor relativamente bajo de 1,4.
- Los otros valores se fijan de la siguiente forma utilizando valores convencionales:
- $Z_{\beta} = 0,84$. Eso corresponde a un poder estadístico del 80%;
- $Z_{\alpha/2} = 1,96$. Eso corresponde a nivel de significación $\alpha = 5\%$.

Utilizando las tablas en Hsieh (1989) se obtiene:

$$n_U = 339$$

En el caso más general, en el cual hay varios factores que pueden influir en la intención migratoria que se quieren tener en cuenta en el modelo, el tamaño de la muestra se tiene que ajustar de la siguiente fórmula:

$$n_M = n_U / (1 - \rho^2)$$

donde ρ es el coeficiente de correlación múltiple entre la variable X y las otras variables independientes.

Utilizando un valor relativamente elevado para ciencias sociales de $\rho = 0,4$ se obtiene:

$$n_M = 339 / (1 - 0,16) = 404.$$

La eficiencia de un muestreo complejo en comparación a la de un muestreo aleatorio simple se define efecto del diseño (“designeffect”), $deff$. Estimar el efecto total de las características del muestreo sobre el $deff$ es complicado en ausencia de un estudio piloto. Utilizaremos un valor de $deff$ de 1,48 que es el valor de $deff$ medio en las encuestas Demographic and Health Surveys (DHS; Aliaga and Ren, 2006).

$$n_{ajustado} = deff * n_M = 1,48 * 404 = 598.$$

A este valor se aplica una corrección que tiene cuenta de la probable no respuesta. Estimando ésta en el 5%, el número final de personas a contactar serán: 623 encuestas.

El tamaño de la muestra anteriormente calculado (623) se tiene que repartir en estos 5 estratos. A cada estrato se asigna una cuota de la muestra total proporcional a su dimensión (% de individuos en el estrato; “repartición proporcional”).

Dentro de cada grupo de estratos se calcula el peso de cada estrato (o sea la proporción de la población de este estrato; tercera columna de la Tabla 3). Luego se multiplica este peso por el valor de la muestra asignada. El número obtenido se redondea por exceso al número entero más próximo. El valor final representa la muestra asignada a cada estrato (última columna de la tabla 2). Se puede notar que a causa del redondeamiento el tamaño final de la muestra ha sido ligeramente aumentado a 625.

CÁLCULO DEL PESO Y MUESTRA DE CADA ESTRATO

ZMV	PARTICIPANTES EN PROYECTOS	POBLACIÓN	PESO	MUESTRA
GT05	202	8654	0.01659	11
GT06	140145	431868	0.82793	516
GT07	274	16863	0.03233	21
GT10	471	36668	0.07030	44
GT11	1043	27573	0.05286	33
Total	142135	521626	1	625

GENERACIÓN DE LA MUESTRA DE CONGLOMERADOS CON EL SOFTWARE ESTADÍSTICO SPSS VERSIÓN 20

Dentro de cada estrato, el muestreo se realiza en dos etapas. En la primera etapa se elige de forma aleatoria una muestra de comunidades. Para la elección de las comunidades se utiliza el **software SPSS, versión 20** y se asigna probabilidades desiguales en función del tamaño (número de habitantes de las comunidades) en la elección, como se recomienda por ejemplo en la literatura (por ejemplo, Levy & Lemeshow, 2013). En el caso de conglomerados con tamaños diferentes, este tipo de muestreo garantiza una eficiencia más elevada.

Posteriormente, dentro de cada comunidad elegida se procederá a la selección aleatoria de una muestra de hogares. El número de hogares a elegir al azar dentro de cada comunidad seleccionada al azar será igual al tamaño final de la muestra asignada a cada estrato (con la repartición proporcional descrita antes) dividido entre el número de comunidades acordado para ese estrato según su peso.

COMUNIDADES IDENTIFICADAS INICIALMENTE DE FORMA ALEATORIA PARA LA REALIZACIÓN DE ENCUESTAS DE ACUERDO AL TAMAÑO DE MUESTRA CALCULADO

<	COMUNIDAD (LUGAR_POBLACIÓN)	MUNICIPIO	POBLACIÓN	PESOS	MUESTRA
GT05	CARACOL	USPANTAN	1115	1	11
Total			1115	1	11
GT06	AP	CONCEPCION HUISTA	854	0,0165	9
GT06	LA VICTORIA	SAN JUAN OSTUNCALCO	5186	0,1001	52
GT06	COMANIBAL	CHICHICASTENANGO	2021	0,0390	20
GT06	ACUL	NEBAJ	3365	0,0649	34
GT06	EL PORVENIR TALQUICHO	SAN LORENZO	1860	0,0359	19

GT06	7A. CALLE 7-02 ZONA 3	MOMOSTENANGO	5117	0,0987	51
GT06	ALDEA BARRANECHÉ	TOTONICAPAN	2250	0,0434	22
GT06	ALDEA BUENA VISTA	SAN BARTOLO AGUAS CALIENTES	1050	0,0203	10
GT06	ALDEA CHAJABAL	SAN ANDRES XECUL	3355	0,0647	33
GT06	ALDEA EL RANCHO	SANTA MARIA CHIQUIMULA	3010	0,0581	30
GT06	ALDEA LOS CIPRESES	MOMOSTENANGO	2738	0,0528	27
GT06	ALDEA NIMASAC	SAN ANDRES XECUL	3743	0,0722	37
GT06	ALDEA TUNAYAC	MOMOSTENANGO	1981	0,0382	20
GT06	CANTON SAN LUIS SIBILA	SANTA LUCIA LA REFORMA	1493	0,0288	15
GT06	CANTON XECAJA	SANTA MARIA CHIQUIMULA	2462	0,0475	25
GT06	CANTON XECAJA, PARAJE XECAMANTUX	SANTA MARIA CHIQUIMULA	852	0,0164	8
GT06	CASERIO CHI-PEREZ, ALDEA BARRANECHÉ	TOTONICAPAN	1286	0,0248	13
GT06	CASERIO PASAC ALDEA SAN VICENTE BUENABAJ	MOMOSTENANGO	1131	0,0218	11
GT06	CASERIO PASUC ALDEA LOS CIPRESES	MOMOSTENANGO	3974	0,0767	40
GT06	PARAJE CHUICAXTUN CANTON CHIYAX	TOTONICAPAN	883	0,0170	9
GT06	PARAJE PASAJOC CANTON MACZUL	TOTONICAPAN	710	0,0137	7
GT06	PARAJE RANCHO CANDELARIA ALDEA PITZAL	MOMOSTENANGO	1033	0,0199	10
GT06	PARAJE TASABALQUIEJ ALDEA RACANA	SANTA MARIA CHIQUIMULA	1464	0,0283	15
Total			51818	1	517
GT07	CHUCHUCA	ZACUALPA	1564	1	21
Total			1564	1	21
GT10	RODEO	CAMOTAN	1148	0,4854	21
GT10	TUTICOPOTE ABAJO	OLOPA	1217	0,5146	23

Total			2365	1	44
GT11	LOS ANGELES LAS PILAS	EL RODEO	899	0,4071	13
GT11	EL CARMEN	SAN PABLO	462	0,2092	7
GT11	EL NARANJO	SAN RAFAEL PIE DE LA CUESTA	847	0,3836	13
Total			2208	1	33
TOTAL					626

COMUNIDADES EN LAS QUE FINALMENTE SE EFECTUARON ENCUESTAS TRAS AJUSTES DEL MUESTREO Y ADAPTACIÓN A LA AUSENCIA DE SUFICIENTES RESPUESTAS EN ALGUNAS DE ELLAS

COMUNIDAD	NÚMERO DE ENCUESTAS REALIZADAS	COMUNIDAD	NÚMERO DE ENCUESTAS REALIZADAS
ALDEA ACUL	35	CASERÍO PASAC	16
ALDEA BARRANECHÉ	27	CASERÍO PASUC	18
ALDEA BUENA VISTA	17	CASERÍO PASUC, ALDEA LOS CIPRESES	17
ALDEA CASA BLANCA	1	CASERÍO PUEBLO VIEJO	3
ALDEA CHAJABAL	36	CENTRO	10
ALDEA CHUCHUCA	11	CHIMATAT	1
ALDEA EL PORVENIR, TALQUICHO	15	EL CARMEN	7
ALDEA EL RANCHO	32	EL NARANJO	14
ALDEA ILOM	24	EL PROGRESO	1
ALDEA IXLA	8	EL RODEO	21
ALDEA LA VICTORIA	3	LA ESPERANZA, QUETZALTENANGO	1
ALDEA LOS CIPRESES	29	LOS ÁNGELES, LAS PILAS	13
ALDEA NICAJA	1	PACHAJ	2
ALDEA NIMASAC	36	PARAJE CHIAXT	1
ALDEA NUEVA CONCEPCIÓN	2	PARAJE CHOESTANCIA, ALDEA PITZAL	2
ALDEA PITZAL	13	PARAJE CHUCUYUP	1
ALDEA PITZAL, PARAJE CHOESTANCIA	17	PARAJE CHUICAXTUN, CANTÓN CHIYAX	3
ALDEA PUEBLO VIEJO	1	PARAJE COXOM, ALDEA CHIMENTE	12
ALDEA SAN ANTONIO POLOGUA	1	PARAJE PACHUM, ALDEA XESANA	25
ALDEA SANTA ANA	1	PARAJE PASAJOC, CANTÓN MACZUL	8
ALDEA SHESHICAP	1	PARAJE RANCHO CANDELARIA, ALDEA PITZAL	11
ALDEA SOTZIL	21	PARAJE TASABALQUIEJ, ALDEA RACANA	16

ALDEA TUNAYAC	26	PARAJE TZANCORRAL, CANTÓN GUALTUX	7
ALDEA XEQUENAYA	3	PARAJE XEJUP, ALDEA LOS CIPRESES	1
AP	10	PICHILUTE	1
BARRANIQUE	1	POLOGUA	1
BARRIO PATZITE	1	RÍO COMANIBAL	12
BARRIO SANTA ANA	7	SAN JUAN OSTUNCALCO	10
BARRIO SANTA CATARINA	3	SAN LORENZO	1
BARRIO SANTA ISABEL	13	SANTA LUCIA LA REFORMA	3
CANTÓN CHIYAX	3	SANTA MARIA CHIQUIMULA	8
CANTÓN PATSAN	1	SECTOR SANTA CATARINA ZONA 3	1
CANTÓN SAN LUIS SIBILA	23	TOTONICAPÁN	4
CANTON XECAJA	41	TUTICOPOTE ABAJO	14
CANTÓN XECAJA, PARAJE XECAMANTUX	7	TUTICOPOTE ABAJO, CASERÍO LOS GARCÍA	8
CARACOL	11	ZONA 3	2
CASERÍO CHIPEREZ, ALDEA BARRANCHE	19		