

Patrones de autoidentificación etnorracial de la población indígena en las encuestas de hogares en el Perú

Martín Moreno*

The Pennsylvania State University, Estados Unidos

* Martín Moreno es candidato a doctor en Sociología y Demografía por la Universidad Estatal de Pensilvania en Estados Unidos donde obtuvo su maestría en Sociología. Asimismo, es bachiller en Sociología por la Universidad Nacional Mayor de San Marcos de Perú. Este estudio fue posible gracias a un financiamiento de la Fundación Ford, a través de un proyecto de GRADE.

Patrones de autoidentificación etnoracial de la población indígena en las encuestas de hogares en el Perú

RESUMEN

Este estudio define y caracteriza los patrones de autoidentificación etnoracial de la población peruana haciendo uso de la Encuesta Nacional de Hogares (ENAHOG) 2012. Para ello utilizamos un módulo con preguntas recientemente incorporado a la encuesta, el cual nos permite aproximarnos empíricamente a la noción de identidad integrando las dimensiones de la lengua o idioma materno, el autorreconocimiento y la noción de pueblo indígena. Los resultados se basan en un análisis de clases latentes usando simultáneamente la información de estas preguntas, las cuales permiten identificar hasta cuatro posibles identidades etnoraciales entrelazadas en el autorreconocimiento con ciertos antepasados y costumbres, así como el que está basado en la lengua. Así, con las herramientas metodológicas usadas, y con la actual modalidad de recolección de datos, no se encuentra evidencia que indique que las identidades étnicas son múltiples y fluidas. Las identidades étnicas muestran cierta heterogeneidad en términos de las dimensiones que las componen.

Palabras clave: identidades étnicas, población indígena, clases latentes

Ethno-racial self-identification patterns of indigenous population in household surveys in Peru

ABSTRACT

This study defines and characterizes ethno-racial self-identification patterns of the Peruvian population employing the Encuesta Nacional de Hogares (ENAHOG) 2012. For this purpose, we have employed a module with questions recently added to this survey, in order to approach empirically to the notion of identity, combining the dimensions of native language, self-acknowledgment and the notion of indigenous population. The results are based on a latent classes analysis employing, at the same time, the information in this questions in order to identify four possible ethno-racial identities. These identities are intertwined in the self-acknowledgment process with certain ancestors and customs, and also the one based on language. With the methodology and the data collected, we haven't found evidence that point that ethnic identities are multiple and fluent. The ethnic identities show certain heterogeneity in terms of the dimensions they are composed.

Keywords: ethnic identities, indigenous population, latent classes

INTRODUCCIÓN

El último censo de población y vivienda realizado en Perú el año 2007 estimó en 3 919 314 el número de personas identificadas como población de origen indígena. Esta cifra comprende a todas las personas de cinco años o más que aprendieron al menos un idioma o lengua materna en su niñez (entre las que figuran el quechua, aimara, asháninka y otras lenguas nativas), segmento que representa al 15,9% de la población (Instituto Nacional de Estadística e Informática, 2008). En la actualidad este es el único indicador oficial que permite estimar el tamaño de la población indígena en el país. Sin embargo, la reciente discusión sobre el uso de nuevas alternativas para clasificar a las poblaciones según su origen étnico y racial sugiere la necesidad de considerar diferentes formas de clasificación que recojan la multidimensionalidad de la noción de etnia y raza (Valdivia 2011, 2012) y que superen los criterios esencialistas (Balarín, 2014).

La combinación de métodos con criterios objetivos y subjetivos ha logrado cada vez mayor validez, y dentro de ellos, el uso de la autoadscripción constituye el método en la comunidad académica y la práctica pública para la identificación de una población desde una perspectiva étnico-racial. Las investigaciones realizadas por Del Popolo y colegas para América Latina documentan un cambio en la modalidad predominante de reconocimiento de las poblaciones de origen indígena y afrodescendientes. Así, varias agencias estadísticas en la región han ido ajustando sus censos, de modo que la autoidentificación ha ganado progresivamente mayor espacio y legitimidad para reconocer a las poblaciones indígenas y afrodescendientes en cada país (Del Popolo, 2008; Del Popolo, Oyarce, Schkolnik y Velasco, 2009). En el Perú, se observan cambios bastante acelerados desde inicios de la primera década del siglo XXI, que es cuando se fueron introduciendo cambios incipientes en la recolección de la etnicidad de la población (Llorens, 2002). Desde el Instituto Nacional de Estadística —la agencia oficial de censos en el país—, así como desde otras instituciones, aún no se ha logrado alcanzar un consenso con respecto al formato de preguntas más adecuado que permita integrar apropiadamente las preocupaciones sobre la visibilización de los diversos grupos etnoraciales (Valdivia, 2011).

Mediante este estudio buscamos contribuir a la discusión sobre la validación empírica de algunas medidas utilizadas para la recolección de información étnico-racial en el país. Específicamente, en este estudio buscamos validar si las herramientas metodológicas propuestas para recolectar la identificación de la población y de los grupos de origen indígena y afrodescendiente constituyen una forma válida que permita mejorar los procesos de visibilización estadística de estos grupos.

El estudio está compuesto de seis secciones, incluyendo esta introducción. En la segunda sección presentamos y discutimos algunos de los hallazgos de la literatura

sobre autoidentificación y derivamos algunas hipótesis de trabajo a partir de dicha literatura. En la tercera sección describimos los métodos usados para el análisis, así como el origen de la base de datos utilizada. La cuarta sección presenta resultados de corte descriptivo, para luego realizar un modelamiento de clases latentes y un análisis multivariado que busca explicar la existencia de asociaciones entre determinadas características. Finalizamos con las conclusiones y las referencias bibliográficas respectivas.

1. LITERATURA SOBRE LA AUTOIDENTIFICACIÓN A TRAVÉS DE ENCUESTAS

La discusión académica sobre cómo operan los procesos de autoadscripción, sobre su naturaleza rígida/flexible, y sobre cómo medirlos a través de encuestas, es extensa. En gran medida, se ha buscado tratar de entender en qué forma la autoadscripción puede ser explicada a partir de su relación de dependencia con otras dimensiones. Esta posición está en contraposición a una visión esencialista y autónoma, donde la identidad étnica y racial no se construye si no que viene dada, visión usualmente asociada con las teorías de origen primordialista (Moreno y Oropesa, 2012). Los procesos de autoidentificación, según esta nueva posición, se asumen casi siempre con respecto a otras características. Esta relación de dependencia permite proponer que la construcción de la identidad sea un proceso maleable y capaz de ser recreado constantemente en función de las circunstancias en las que se desarrolle (Moreno y Oropesa, 2012).

Schkolnik (2009) propone un mínimo de dimensiones a partir de las que se pueda operacionalizar la pertenencia a un grupo étnico. La primera dimensión la constituye el origen común o la presencia de los ancestros comunes. Así, la vinculación a una historia y un pasado común contribuye a definir un sentido común de pertenencia a un determinado grupo étnico. La segunda dimensión corresponde a la cultura de origen, la que a su vez incluye una serie de marcadores, como por ejemplo el idioma, la religión y sus prácticas cotidianas, el estilo de vida —siendo este un concepto bastante amplio—, así como las formas de organización social y política. La tercera dimensión es la denominada «conciencia étnica» e implica la capacidad de autorreconocerse en una identidad étnica específica, como parte de un pueblo diferenciado de los demás y con derechos generados por la pertenencia a aquel. La última dimensión que la autora desarrolla es la denominada «territorialidad», y corresponde a la ocupación geográfica de un determinado espacio dentro del cual es posible encontrar prácticas sociales comunes, tanto cotidianas como relacionadas con la herencia ancestral de los grupos que ocupan dichos territorios (Schkolnik, 2009, p. 67).

La propuesta de esta autora reconoce que la medición de la dimensión relacionada con los antepasados comunes no está claramente delimitada, y por ende,

genera problemas de medición. Por otro lado, no queda claro que un acercamiento empírico a las otras dimensiones esté exento de problemas. Sin embargo, los problemas de medición parecen no limitarse necesariamente a aquella única dimensión. Si tomamos en cuenta que la dimensión de la conciencia étnica, así como la dimensión cultural, comparten rasgos, se hace difícil establecer con claridad qué elementos podrían ser usados para separarlas. Resulta entonces que la territorialidad figura como la única dimensión distinguible de las otras y que puede ser empíricamente anclada.

Schkolnik también enfatiza que cualquier aproximación empírica a la etnicidad presupone la exclusión de ciertas categorías que denominaremos «contaminantes». El ejemplo aquí se aplica a categorías como *mestizo* o *blanco*, las cuales pueden atraer respuestas de quienes no desean hacer explícita una identidad indígena o, eventualmente, esconderla (Schkolnik, 2009).

Una preocupación que emerge de una conceptualización de la etnicidad a partir de estas dimensiones es que, a pesar de que se recoge la perspectiva subjetiva y el derecho de los individuos y las comunidades a identificarse con una determinada etnicidad, se descarta a la vez una posible *multidimensionalidad simultánea* de los procesos de autorreconocimiento. Es decir, se vuelve a cierta forma de esencialismo, donde se asume que ciertas características y marcadores son vitales para definir la pertenencia o exclusión de cierto grupo: así, se impone cierta rigidez conceptual y se cierra la posibilidad a que las fronteras resulten más porosas de lo que pensamos.

Las categorías oficiales usadas por el Estado y sus instituciones contribuyen a delimitar divisiones étnicas y fronteras sociales que ciertos grupos pueden aceptar con facilidad o no. La dificultad de adscribirse con una determinada división y clasificación es el resultado de una fricción entre la población que podría aceptar dicha división y su rechazo a ella por la connotación que conlleva y las consecuencias que implica (Wimmer, 2007). Así, la aceptación o rechazo hacia el uso de ciertas categorías responde a una estrategia que es a su vez el reflejo de intentar encajar en un determinado orden o jerarquía étnica.

Una explicación alternativa para entender el proceso de construcción de las identidades étnicas es que estas son en realidad entidades moldeables, y que es posible expandir la idea de que existen múltiples dimensiones de dicha identidad que se superponen una a la otra, en algunos casos de manera complementaria, en otros de modo redundante. La noción de procesos de identificación étnicos flexibles no es nueva y ya ha sido analizada empíricamente por algunos autores en el caso del Perú. Por ejemplo, Moreno y Oropesa sostienen que, entre personas residentes en zonas urbanas del Perú, las personas de origen indígena son más propensas a cruzar la frontera étnica y usar la categoría «mestizo» bajo ciertas condiciones (como por ejemplo, tener padres monolingües, o al menos que uno de ellos hable

una lengua indígena). Recientemente, Planas, Cruzado, Middelkoop, Cárcamo y Richters (2013) retoman el tema desde una perspectiva de autoidentificación con múltiples identidades entre mujeres de origen indígena. Las autoras proponen la idea de navegación como forma de retar al discurso de fronteras étnicas rígidas, aunque lo condicionan a la existencia de ciertas condiciones para que esto ocurra. Por ejemplo, las experiencias de migración urbana, de exposición permanente a una lengua mayoritaria, o el acceso a la educación generan nuevas oportunidades para «de-indianizarse» (Moreno y Oropesa, 2012).

2. IMPLICACIONES DE DOS MODELOS DE AUTOIDENTIFICACIÓN ÉTNICA

Las perspectivas conceptuales presentadas previamente sugieren la existencia de dos posibles resultados empíricos en términos de autoidentificación étnica. Por un lado, la primera perspectiva nos sugiere que, en el plano individual, la autoidentificación conlleva a la elección de una identidad consistente mediante categorías claramente excluyentes, y que esa elección se reiteraría a través de las diferentes dimensiones con las que pueda ser medida. Así, la elección de una categoría étnica asociada a la dimensión cultural sugiere que se mantendrá un uso consistente con una categoría similar dentro de la dimensión de ancestralidad. La hipótesis que se deriva de este mecanismo es que la elección de una categoría asociada a una identidad étnica particular conlleva a un incremento en la probabilidad de que se escojan categorías similares en otras de las dimensiones relacionadas con la etnicidad.

El otro modelo que se deriva de la conceptualización de la etnicidad como un proceso flexible entiende que es posible encontrar una menor concordancia en los patrones de respuestas a las preguntas sobre autoidentificación. En este modelo, sería posible esperar que no exista necesariamente consistencia en las respuestas en cada dimensión y que, por el contrario, los encuestados puedan expresar su identificación étnica dentro de una dimensión con una elección distinta en otra dimensión. La hipótesis que surge es que la ausencia de consistencia étnica refleja la existencia de una multiplicidad de identidades étnicas sin mayor dominio de unas sobre otras.

Nuestra hipótesis de trabajo es que, si efectivamente los tres criterios que usamos son coincidentes entre sí al medir la etnicidad-racialidad de la población, entonces identificaremos grupos de individuos con patrones de respuesta similares y con una alta asociación entre sí. Por otro lado, si una o varias dimensiones capturan aspectos de la etnoracialidad que no se entrelazan entre sí, entonces esperamos una mayor dispersión en el patrón de respuestas, lo que refleja una mayor fragmentación en los grupos.

La identificación de los patrones de respuestas, y cómo estos se relacionan con identidades fluidas o identidad fijas y consistentes, no puede omitir el estudio de

los posibles mecanismos que expliquen cómo se produce dicha identificación. Proponemos aquí tres posibles mecanismos que expliquen los patrones de identificación y la existencia de determinadas identidades. En primer lugar, la idea de de-indigenización supone que existen ciertas características básicas de los individuos que explican su mayor propensión a identificarse con categorías étnicas no indígenas. El segundo mecanismo es el de la transmisión intergeneracional. Aquí importa entender en qué medida la elección de ciertos patrones étnicos responden a la transmisión de cierta identidad entre generaciones. Finalmente, tenemos el rol que juega el contexto donde se manifiestan: la mayor concentración de un determinado grupo étnico estaría asociada con la mayor propensión a escoger un grupo étnico similar.

3. OBJETIVOS

El estudio busca responder a una serie de preguntas; algunas con implicancias de política, otras referidas a las aplicaciones empíricas. Respecto de las implicancias de política, buscamos indagar cómo contribuye el instrumento propuesto para la recolección de datos étnicos a determinar la existencia de identidades etnorraciales en el Perú. En la forma propuesta actualmente por el INEI, el instrumento aplicado en la Encuesta Nacional de Hogares permite mejorar la identificación y la visibilización de grupos etnorraciales que de otra forma no tendrían mayor visibilidad. Pero queda pendiente establecer si construir una medida sintética que refleje la simultaneidad de diferentes criterios de identificación, ancestralidad, idioma y conciencia étnica contribuye a mejorar el proceso de identificación de grupos étnicos en el Perú. La idea es ver si las preguntas aplicadas logran mejorar o aclarar el proceso de identificación de ciertos grupos étnicos. El segundo objetivo es identificar algunos de los mecanismos que contribuyen a la formación de ciertos patrones de respuesta.

Entonces, ¿cuál es la mejor medida para introducir en un instrumento masivo de recolección de datos? ¿Cuál debería ser el criterio para descartar —o, por el contrario, privilegiar— la inclusión de ciertas preguntas? Aquí proponemos que lo que deberíamos hacer es seleccionar las preguntas que maximizan la visibilización de las poblaciones indígenas y afrodescendientes, es decir, que identifiquen a la mayor proporción de la población indígena, pero sin recurrir a la construcción de una medida que infle artificialmente los grupos poblacionales o que carezca de validez.

En este estudio proponemos que la autoidentificación basada en la multiplicidad de criterios representa una herramienta más completa en términos de ofrecer mayor libertad. Ello no niega la posibilidad de complejizar el análisis a través de la multiplicidad de ancestros.

4. MÉTODOS Y DATOS

4.1. Estrategia analítica

Para establecer si los patrones de respuesta a las preguntas de etnicidad responden a un modelo de clasificación de identidad única o a un modelo de identidades múltiples, el análisis de los datos se organiza de la siguiente manera. En una primera etapa, describimos la distribución de las respuestas a la pregunta de autoidentificación étnico-racial con opciones múltiples para luego identificar los patrones de respuesta existentes. Asimismo, usando una serie de análisis bivariados, exploramos algunos de los mecanismos que median los patrones de respuestas en las preguntas de autoidentificación. Luego empleamos la técnica de análisis de clases latentes para identificar si existen grupos de la población que pueden ser agrupados en función de los patrones de respuestas observados. Finalizamos con la caracterización de los grupos étnicos encontrados; aplicamos un modelo de regresión multinomial para identificar las características que se encuentran más asociadas con las clases latentes identificadas.

4.2. Datos y variables

Los análisis en este estudio se basan en los datos de la Encuesta Nacional de Hogares (ENAH0) del año 2012, conducida por el Instituto Nacional de Estadística (INEI) del Perú. Específicamente, este estudio hace uso de las preguntas de identificación étnica y racial (antepasados/cultura y pertenencia a pueblos indígenas), además de la pregunta sobre lengua materna de los encuestados.

La ENAH0 recolecta información demográfica, económica, educativa y laboral de todos los miembros del hogar, y tiene representatividad nacional. El diseño muestral rotativo de la encuesta permite recolectar información durante todo el año de manera continua. La muestra analítica usada para este estudio comprende 73 329 individuos de catorce años de edad o más que respondieron alguna de las preguntas que miden la identificación étnica y racial. Estos individuos representan a 25 091 hogares localizados en 24 624 viviendas particulares.

4.3. Variables utilizadas para el análisis

Medidas de la autoidentificación étnica y racial

Para analizar los patrones de autoidentificación etnoracial se utilizaron tres preguntas: el idioma o lengua materna del entrevistado (módulo 300, pregunta 300A), la autoidentificación con un grupo con el que comparte antepasados/cultura similar (módulo 500, pregunta 558C) y la pertenencia a un pueblo indígena (módulo 500, pregunta 558D). La transcripción literal de las preguntas usadas se puede encontrar en el anexo 1.

La primera pregunta, idioma o lengua materna del entrevistado, sigue un formato estándar usado en encuestas previas, así como en la recolección del último censo del año 2007. Las categorías consideradas en la ENAHO 2012 son quechua, aimara, otra lengua nativa —con posibilidad de especificarla—, castellano, inglés, portugués, otra lengua extranjera —es posible especificar— y si la persona es sordomuda. La segunda pregunta busca identificar la identidad con la cual la persona encuestada guarda mayor afinidad. Específicamente, el fraseo de la pregunta considera simultáneamente a los antepasados y a las costumbres como puntos de referencia, y propone como alternativas de respuesta a las categorías quechua, aimara, nativo o indígena de la Amazonía, negro/mulato/zambo/afroperuano, blanco, mestizo, otro y no sabe. Finalmente, la tercera pregunta indaga si el encuestado declara pertenecer o considerarse parte de un pueblo indígena. Las respuestas a la tercera pregunta son del tipo afirmación-negación, incluyendo la posibilidad de declarar que no sabe¹.

La ENAHO 2012 introdujo dos variaciones relevantes respecto de encuestas previas sobre cómo se recolectaba información étnico-racial. La primera es que la pregunta usada regularmente para recolectar la autoidentificación étnica-racial según antepasados/costumbres fue incluida en el módulo de empleo y salud, con lo que se logró mayor cobertura al aplicarse a toda la población mayor de catorce años de edad. Anteriormente, la pregunta era parte del módulo sobre gobernabilidad, y era aplicada únicamente al jefe del hogar y a su cónyuge. La segunda variante es que la pregunta sobre el sentido de pertenencia a un pueblo indígena es nueva, siendo incorporada en el módulo de empleo y salud. Ambas preguntas fueron ubicadas al final del módulo y preguntadas de manera secuencial.

Variables asociadas

En el análisis de los determinantes de la elección de una identidad étnica o racial, la literatura especializada sugiere la inclusión de una serie de características en los ámbitos individual, familiar y de contexto, para entender en qué medida las decisiones de identificación están condicionadas por dichas características o si son completamente independientes de ellas.

Las características demográficas más comúnmente usadas son el sexo de la persona, la edad o cohorte, el estado civil y la exposición a procesos migratorios. En términos familiares, se trata de indagar por los antecedentes socioeconómicos de los padres, en tanto ellos pueden contribuir a entender los procesos de autoadscripción. Aquí usaremos la educación de los padres y también consideraremos la educación del individuo. Finalmente, exploraremos el rol de variables de contexto

¹ La encuesta también indagó sobre el pueblo indígena específico al que declara pertenecer el entrevistado. Sin embargo, la base de datos a la que tuvimos acceso no contenía información válida en los campos (variables) asociados a dicha pregunta, por lo cual fue excluida del análisis.

como las variables de composición étnica de las zonas de residencia de los encuestados. Incluiremos información secundaria del distrito de residencia del encuestado, en particular, la proporción de la población en el distrito que habla una lengua nativa según el censo de 2007 y el número absoluto de comunidades nativas.

4.4. Métodos analíticos: análisis de clases latentes

Para identificar a partir de los patrones de respuestas a los grupos étnicos existentes empíricamente, usamos análisis de clases latentes o, como se le conoce en inglés, *Latent Class Analysis*. El análisis de clases latentes es un método de análisis estadístico que permite identificar grupos de individuos que, al interior de cada grupo, son altamente similares entre sí, pero que entre ellos son mutuamente excluyentes. Las comunalidades y divergencias entre los grupos se estiman basándose en los patrones de respuesta a una serie de preguntas de tipo categórico (Lanza, Collins, Lemmon y Schafer, 2007). Los patrones de respuesta que interesa identificar corresponden a las alternativas seleccionadas en las tres preguntas sobre identificación étnica y racial. Así, construimos una pregunta con múltiples indicadores de respuestas que el análisis de clases latentes utiliza para identificar a los individuos que muestran una similar propensión estadística a responder usando un patrón de respuesta determinado, para luego agruparlos en clases latentes.

El supuesto de este método es que la muestra bajo estudio proviene de una población que consta de varios grupos (K grupos) que no pueden ser observados directamente. Sin embargo, es posible inferir la pertenencia o no a los grupos a través de ciertos indicadores observados usualmente a través de una encuesta. Se considera que cada observación en la muestra ha sido escogida de una de las K subpoblaciones o clases latentes, y K podría conocerse o no. Así, cuando los individuos responden a las preguntas sobre su identificación con un grupo etnorracial de manera consistente, asumimos que existe un modelo de medición que puede ser estimado para reflejar dicha relación. Al asumir que los grupos raciales se manifiestan de manera heterogénea, podemos esperar que un modelo de varias clases sea el más adecuado. El análisis de clases latentes, entonces, forma parte de los llamados modelos de heterogeneidad no observada, o modelos de mezcla finita. Cada individuo es clasificado de manera excluyente y exclusiva en un solo grupo o clase latente, y para ello se usa la probabilidad posterior calculada para cada individuo (Muthen, 2001, p. 7). En la medida que nuestro interés es identificar cuántos y qué características tienen los grupos etnorraciales, estimaremos los modelos sin covariantes.

5. RESULTADOS

5.1. Características de la muestra analítica

En la tabla 1 reportamos las principales características de la muestra analítica. Esencialmente es una muestra compuesta por más de 70 000 observaciones, con una presencia de mujeres ligeramente mayor que de hombres (51,5% y 48,5%, respectivamente), donde tres de cada cuatro entrevistados viven en zonas urbanas, y según la edad del entrevistado, muestran una concentración ligeramente mayor de personas entre 20 y 29 años de edad, cuyo grupo representa aproximadamente 20,5% del total de la población analizada. Observamos una fuerte concentración de población que ha logrado alcanzar educación secundaria. Específicamente, 43% de la muestra ha logrado acceder a educación secundaria. Si a ello le sumamos los que lograron acceder a educación superior, el logro educativo de educación secundaria alcanza al 70% de la muestra.

Tabla 1. Datos generales de la muestra desagregados por área

	Área			Área			Área		
	Urbana Casos	Rural Casos	Total Casos	Urbana % Ver.	Rural % Ver.	Total % Ver.	Urbana % Hor.	Rural % Hor.	Total % Hor.
Total	46 194	27 135	73 329	100,0	100,0	100,0	76,6	23,4	100,0
Sexo									
Hombre	22 249	13 634	35 883	48,0	50,2	48,5	75,8	24,2	100,0
Mujer	23 945	13 501	37 446	52,0	49,8	51,5	77,4	22,6	100,0
Edad									
14-	7 436	5 434	12 870	15,3	20,2	16,4	71,2	28,8	100,0
20-	9 524	4 392	13 916	20,5	16,1	19,5	80,8	19,2	100,0
30-	7 960	4 347	12 307	16,9	15,8	16,7	77,9	22,1	100,0
40-	7 750	4 299	12 049	16,7	15,7	16,5	77,8	22,2	100,0
50-	6 215	3 414	9 629	13,8	12,6	13,5	78,2	21,8	100,0
60-	7 309	5 249	12 558	16,8	19,6	17,4	73,7	26,3	100,0
Máximo nivel educativo alcanzado									
Primaria	10 181	15 327	25 508	20,9	57,7	29,5	54,3	45,7	100,0
Secundaria	20 624	9 738	30 362	45,0	35,3	42,7	80,7	19,3	100,0
Terciaria	15 350	2 065	17 415	34,1	6,9	27,8	94,2	5,8	100,0

Nota: Los porcentajes reportados en la tabla están ajustados para reflejar los pesos poblacionales del diseño muestral de la encuesta.

Fuente: Encuesta Nacional de Hogares 2012. Elaboración propia.

5.2. Análisis descriptivo

En esta sección reportamos los patrones de respuesta encontrados en cada pregunta por separado.

Dependiendo de la forma de medición/clasificación, el tamaño de la población indígena fluctúa entre 17% y 28%. Si consideramos la pregunta sobre idioma/lengua materna (ver tabla 2) encontramos que aproximadamente 19,2% de las personas encuestadas reporta tener como lengua materna una lengua indígena y la mayor parte de ellas es quechuahablante (16,7%)². Según la lengua indígena en la que sea clasificada la población, observamos algunas diferencias relevantes en términos contextuales. La población quechuahablante se concentra en la sierra sur y sierra centro (38% y 32% respectivamente), así como en Lima Metropolitana (20%), mientras que entre la población aimara, tres de cada cuatro personas viven en la sierra sur. La población con alguna lengua nativa exhibe un grado de concentración geográfica mucho mayor: casi la totalidad reside en la selva (96%), mayormente en zonas rurales. Consecuentemente es el grupo de población que muestra la tasa de migración interdistrital más baja respecto de los otros grupos de hablantes de lenguas nativas (77% de las personas reporta vivir en el mismo distrito donde nacieron, mientras que en los otros grupos este porcentaje es de 55%).

En términos demográficos, se observan algunas diferencias importantes en la distribución de edades entre los quechua y aimarahablantes respecto de los hablantes de otras lenguas nativas³. Así, los hablantes de otras lenguas nativas muestran una mayor concentración de población joven menor de treinta años: aproximadamente 40% de la población en este grupo no excede esta edad, mientras que en los otros dos grupos de hablantes nativos los grupos más jóvenes representan un 16-21% de la población. Esta diferencia indica una dinámica distinta entre ambos grupos, pues en los quechuas y aimaras, el uso del idioma se concentra entre las cohortes más viejas, mientras que entre los hablantes de otras lenguas nativas el uso resalta entre los más jóvenes.

² Según el censo de 2007, el porcentaje de personas de catorce años a más que son identificadas como indígenas es de 16,7% (cálculo propio a partir de datos obtenidos en línea usando el sistema de consulta de datos censales Redatam). Véase: <http://iinei.inei.gob.pe/iinei/RedatamCpv2007.asp>. La diferencia con el porcentaje reportado en este estudio podría ser explicada porque el diseño muestral de la ENAHO excluye de su cobertura centros poblados de difícil acceso, y muy probablemente tienda a incluir mayoritariamente poblaciones de origen indígena que sí son visitadas en el censo.

³ Estos resultados deben tomarse con cierta cautela, considerando el tamaño de la población en este grupo y el alto nivel de variabilidad existente. Sin embargo, estimaciones auxiliares usando el censo de 2007 sugieren que el patrón encontrado en esta encuesta estaría dentro de lo esperado.

Tabla 2. Idioma o lengua materna que aprendió en su niñez

	Observaciones	Porcentaje
Quechua	14 259	16,7
Aimara	1 746	2,0
Otra lengua nativa	975	0,5
Castellano	56 059	80,5
Inglés	61	0,1
Portugués	49	0,1
Otra lengua extranjera	22	0,1
Es sordomudo	110	0,1
Total	73 281	100,0

Nota: Excluye 48 observaciones (0,1%) con respuesta omitida
Los porcentajes están ajustados por los pesos muestrales.

Fuente: Encuesta Nacional de Hogares 2012. Elaboración propia.

La distribución de las respuestas a la pregunta sobre antepasados y costumbres muestra resultados más prometedores, en cuanto identifica a una población de origen indígena más grande. Así, 28% de la población mayor de catorce años se identifica con alguna de las categorías etnorraciales indígenas. Si añadimos a la población afrodescendiente, el porcentaje asciende a 30%. Concentrándonos en las identidades indígenas, encontramos que los patrones de distribución geográfica son similares a los observados en términos de la lengua materna. Sin embargo, un aspecto interesante es que, en el caso de quienes se identifican como nativos o indígenas de la Amazonía, la pregunta logra recuperar a un importante porcentaje de personas residentes en Lima Metropolitana. De manera similar, el porcentaje de concentración geográfica de quienes se identifican como aimaras tiende a redistribuirse. La distribución según edades es homogénea entre quienes se identifican como quechuas y aimaras, mientras que entre quienes se autoidentifican como nativos de la selva la distribución es dominada por la población de menor edad. La tasa de no respuesta a esta pregunta es relativamente baja —aproximadamente 5%— y aparentemente está relacionada con el nivel educativo (38% de los que no responden han alcanzado educación primaria, en comparación con el 30% de la muestra que solo alcanzó primaria) y con la residencia en ciertas zonas del país (la población de la sierra norte y de Lima Metropolitana está ligeramente sobrerrepresentada).

Tabla 3. Por sus antepasados y según sus costumbres, ¿cómo se considera usted?

	Observaciones	Porcentaje
Quechua	19 411	23,6
Aimara	2 595	2,8
Nativo o indígena de la Amazonía	2 173	1,7
Negro/mulato/zambo/afroperuano	1 172	1,9
Blanco	3 350	5,8
Mestizo	37 992	54,3
Otro	3 321	5,3
No sabe	3 233	4,7
Total	73 247	100,0

Nota: Excluye 82 observaciones (0,2%) con respuesta omitida.
Los porcentajes están ajustados por los pesos muestrales.

Fuente: Encuesta Nacional de Hogares 2012. Elaboración propia.

En la pregunta sobre pertenencia a los pueblos indígenas, el porcentaje de respuestas afirmativas alcanza 16,5%. En general, no se observan diferencias importantes en términos sociodemográficos, es decir, las características de quienes se reconocen como miembros de un pueblo indígena se encuentran distribuidas de manera similar respecto de quienes no se identifican como tales. En algunas áreas del país (costa sur, sierra centro y sur) es mayor la probabilidad de encontrar personas que se consideren como parte de un pueblo indígena respecto del resto del país. La tasa de no respuesta es del 6%, y salvo el nivel educativo —es más probable que la población con menor educación no responda esta pregunta— no se observan mayores diferencias con el resto de la población.

Tabla 4. ¿Usted pertenece o se considera parte de un pueblo indígena?

	Observaciones	Porcentaje
Sí	12 421	16,5
No	56 318	77,0
No sabe	4 508	6,6
Total	73 247	100,0

Nota: Excluye 82 observaciones (0,1%) con respuesta omitida.
Los porcentajes están ajustados por los pesos muestrales.

Fuente: Encuesta Nacional de Hogares 2012. Elaboración propia.

Así, las respuestas a las tres formas de autoidentificación evaluadas ofrecen tres resultados distintos y es la pregunta de identificación histórico-cultural (según los antepasados) la que permite identificar una proporción más alta de población de origen indígena: 28%. En el otro extremo, la pregunta sobre pertenencia explícita a pueblos indígenas es la que muestra la menor probabilidad de identificar a una población de origen indígena, si bien son pocas las características que explican las diferencias. Estos resultados reflejan lo que investigaciones en otros contextos señalan: el tamaño de la población indígena está altamente condicionado a los criterios que se usen para definir quiénes forman parte de ella y a los métodos usados para identificarla. Sin embargo, puesto que varias de las categorías y las etiquetas etnoraciales contenidas en estas preguntas capturan dimensiones de la identidad etnoracial que se superponen entre sí, es necesario explorar cuál es el grado de concordancia en los patrones de respuesta.

5.3. Identificación de patrones de respuesta comunes

En la tabla 5 presentamos una matriz con los niveles de concordancia entre las tres modalidades de identificación separadas según identidad indígena (debajo de la diagonal) e identidad mestiza (encima de la diagonal). Así, las personas clasificadas como indígenas usando la pregunta de etnorraza, es decir, las que respondieron algunas de las tres categorías (quechua, aimara, nativo de la Amazonía), también respondieron alguna categoría relacionada con un idioma indígena (quechua, aimara o lengua nativa); estas personas obtuvieron un nivel de concordancia de 17%. El nivel de concordancia es consistentemente más alto en zonas rurales para aquellos con una identidad indígena. En contraste, el nivel de concordancia en las respuestas relacionadas con una identidad mestiza es mucho más alto: 52% de los que respondieron la categoría mestizo en la pregunta sobre ancestros/antepasados reportaron que el castellano es su lengua materna⁴, y el resultado es algo más alto en zonas urbanas (58%). Si sumamos los porcentajes de concordancia entre las tres dimensiones que definen las identidades etnoraciales, obtenemos el nivel de concordancia total. Los resultados confirman un nivel de concordancia más bajo para aquellos que reportan una identidad indígena respecto de quienes reportan una identidad mestiza.

⁴ Utilizamos la categoría *castellano* como indicador de mestizo basándonos en la idea de que los procesos de asimilación cultural conllevan al uso de marcadores que diluyen una identidad nativa fuerte.

Tabla 5. Porcentaje de concordancia entre distintos indicadores de etnicidad

Total

Etnorraza	52,2	6,8
16,5	Idioma	11,2
7,8	5,2	Pueblo indígena

Área urbana

Etnorraza	58,6	7,7
10,0	Idioma	12,3
6,8	4,0	Pueblo indígena

Área rural

Etnorraza	31,3	4,0
37,7	Idioma	7,7
11,2	9,4	Pueblo indígena

Nota: En cada uno de los paneles, las celdas de la matriz ubicadas encima de la diagonal de cada panel indican el porcentaje de concordancia entre cada par de entradas al responder alguna de las siguientes categorías (etnorraza = mestizo; idioma = castellano; pueblo indígena = no). Del mismo modo, las celdas ubicadas debajo de la diagonal de cada matriz representan la concordancia en la identidad indígena (etnorraza = quechua, aimara o nativo de la Amazonía; idioma = quechua, aimara o lengua nativa; pueblo indígena = sí). Por ejemplo, en el panel que corresponde al área rural, 37,7% de los encuestados que dijeron identificarse como quechuas, aimaras o nativos de la Amazonía en la pregunta de etnorraza también lo hicieron en la pregunta de idioma.

Fuente: Encuesta Nacional de Hogares 2012. Elaboración propia.

5.4. Resultados de modelos de clases latentes

Los resultados del análisis de clases latentes están basados en las tres preguntas sobre autoidentificación étnica: antepasados o ancestros, idioma materno y pertenencia a pueblo indígena. Para evitar posibles problemas de convergencia al no poder estimar los resultados con escasa información, tomamos la decisión de reagrupar algunas categorías de respuesta en la pregunta sobre idioma/lengua materna: así, quienes reportaron que hablaban inglés, portugués y otras lenguas extranjeras fueron agrupados en una categoría, además de eliminar las observaciones que respondieron con la categoría sordomudo. Adicionalmente, fueron descartadas del análisis todas las observaciones que tenían como respuesta la categoría «no sabe». Los indicadores de las tres preguntas sobre autoidentificación fueron analizados con un modelo sin covariantes mediante el *plugin LCA* para Stata desarrollado por The Methodology Center de Pennsylvania State University.

Dado que ignoramos *a priori* el número de clases latentes que se ajustan mejor a los datos, estimamos varios modelos con distintos números de clases y escogemos

aquel que ofrezca el mejor ajuste estadístico. Específicamente, usamos como criterio de selección aquel modelo que muestre el menor valor en los estadísticos AIC (*Akaike Information Criterion*) y BIC (*Bayesian Information Criterion* o *Schwarz-Bayesian Information Criterion*)⁵. Otro criterio para preferir un modelo sobre otro es la facilidad de interpretación: por ejemplo, que el tamaño de las clases encontradas sea distinguible y con una probabilidad mayor a cero.

Los resultados que se presentan en la tabla 6 muestran el resultado de aplicar el análisis de clases latentes para identificar la existencia de estructuras latentes comunes usando los patrones de respuesta a las tres preguntas que capturan tres dimensiones de la etnicidad. La aplicación del modelo de clases latentes sugiere que existen cuatro grupos distinguibles y homogéneos entre sí a partir de las respuestas ofrecidas, de los cuales tres grupos identifican claramente identidades de corte étnico, y el restante aparece asociado con la categoría «mestizo».

Las tres clases étnicas identificadas tienen correspondencia con las categorías usadas para describirlas, y llama la atención la claridad con la que dichas identidades se distinguen al interior de cada clase. Así, entre quienes la identidad quechua es saliente (aproximadamente 29% de la población), la autocategorización como quechua según sus costumbres y antepasados es dominante: nueve de cada diez personas la escogen. Respecto del idioma materno dentro de esta clase, nuevamente el quechua aparece como el idioma dominante (61%); aun así, permanece un 39% que reporta al castellano como su idioma materno. Sin embargo, el marcador de pertenencia a un pueblo indígena solo es respondido por uno de cada cuatro dentro de esta clase, lo que sugiere que la noción de pueblo no guarda correspondencia con las otras dimensiones étnicas.

En el caso de la clase aimara, que representa un 3,3% de la población, los patrones de respuesta son similares a los observados dentro de la clase quechua. En su mayoría, la etnicidad se define en función de la presencia de antepasados aimaras (casi 95%) y en función del predominio del idioma aimara (68% de los casos), mientras que encontramos una autopercepción de pertenencia a un pueblo indígena menor a la esperada, aunque ligeramente más alta entre quienes forman esta clase (tres de cada diez) respecto de los de la clase quechua.

La última de las tres clases asociadas a identidades étnicas es aquella que se identifica con una identidad amazónica y representa un 2% de la población. Esta clase guarda similitudes con las otras dos clases previas; sin embargo, es la única de las tres donde la identificación con la categoría «pueblo indígena» sí muestra la intensidad que se esperaría, pues 90% de las personas en esta clase optan por el uso de esta categoría.

⁵ Otras medidas de ajuste usadas son el Chi-Square (ChiSq) o el Likelihood Ratio Chi-Square (Gsq), y son reportadas en el anexo 2.

Tabla 6. Clases latentes en los patrones de respuesta: distribución y patrones de respuesta (en porcentajes)

	Clase con identidad amazónica	Clase con identidad quechua	Clase con identidad aimara	Clase con identidad mestiza
Tamaño (%)	2,0	29,0	3,3	65,7
Por sus antepasados y según sus costumbres, usted se considera:				
1 Quechua	1,0	90,9	0,0	0,1
2 Aimara	0,0	0,0	94,5	0,5
3 Nativo de la Amazonía	84,6	0,1	0,0	2,4
4 Negro	1,0	0,3	0,7	2,3
5 Blanco	0,9	0,3	0,7	6,4
6 Mestizo	4,9	7,7	2,8	82,0
7 Otro	7,5	0,6	1,3	6,3
Idioma o lengua materna que aprendió en su niñez:				
1 Quechua	0,9	61,3	1,2	0,0
2 Aimara	0,0	0,0	68,2	0,0
3 Otra lengua nativa	80,3	0,0	0,0	0,0
4 Castellano	18,7	38,7	30,6	100,0
¿Usted pertenece o se considera parte de un pueblo indígena?				
1 Sí	89,3	26,2	36,8	11,9
2 No	10,7	73,8	63,2	88,1

Fuente: Encuesta Nacional de Hogares 2012. Elaboración propia.

En la tabla 7 presentamos una caracterización de cada una de las clases latentes encontradas, usando como estadísticas descriptivas algunas de las variables disponibles. La distribución de las clases por género es homogénea, con un ligero predominio de las mujeres respecto de los hombres. Con relación a la edad de los informantes, observamos que no existen variaciones y que se encuentran distribuidas de manera similar al total de la población. Las únicas excepciones se darían entre las clases aimara y quechua, donde se observa una incidencia de la población en el intervalo de edad más alto (sesenta años o más) comparado con el resto de los grupos. De manera análoga, entre los indígenas de la Amazonia, la estructura de edades tiende a mostrarse ligeramente más alta entre los segmentos más jóvenes, lo cual podría ser el reflejo de una dinámica demográfica distinta.

Donde comenzamos a observar algunos cambios relevantes en la distribución es en el área y la región de residencia, así como en los patrones migratorios. Por un lado, la identidad amazónica aparece claramente ligada a zonas rurales en la región de la selva y con poca experiencia de migración a diferentes distancias (la tasa de incidencia de migración distrital, provincial o departamental/regional es de 23%, 15% y 11% respectivamente). La identidad quechua se encuentra concentrada en las zonas de la sierra centro y sur, en la selva y en Lima Metropolitana (41%, 31%, 10% y 8% respectivamente), distribuida casi homogéneamente entre zonas urbanas y rurales y con niveles de migración intermedias entre aquellos con identidad amazónica y los aimaras. La clase identificada con los aimaras se encuentra concentrada en zonas urbanas (62%) en la costa y sierra sur (90% entre ambas zonas) y con mucha mayor dinámica migratoria, en particular en distancias largas: la incidencia de la migración interdepartamental es la más alta (38%).

Si asumimos que estas tres características reflejan parcialmente una de las dimensiones omitidas, la denominada «territorialidad», podríamos proponer que es la clase con identidad amazónica aquella donde las dimensiones clásicas de identidad muestran mayor consistencia: si efectivamente la territorialidad guarda relación con el vínculo ancestral y la cercanía con el territorio, entonces aquel grupo sería el más propenso a mostrarla, dado los mucho menores niveles de intensidad de migración fuera del área de nacimiento, así como la fuerte concentración en el espacio geográfico delimitado por la selva (aunque considerando el tema de la espacialidad y las distancias).

Tabla 7. Características sociales y demográficas de cada una de las clases/identidades latentes identificadas

	Identidad amazónica	Identidad quechua	Identidad aimara	Identidad mestiza	Total
	Clase 1	Clase 2	Clase 3	Clase 4	
<i>Sexo</i>					
Hombre	49,6	49,6	49,4	48,7	49,0
Mujer	50,4	50,4	50,6	51,3	51,0
<i>Edad</i>					
14-	17,5	18,1	15,1	17,9	17,8
20-	16,8	16,5	15,9	20,0	18,8
30-	21,9	15,5	16,7	17,2	16,8
40-	16,1	15,8	17,2	16,1	16,0
50-	16,1	15,1	14,6	13,8	14,2
60-	11,7	19,1	20,5	15,1	16,3
<i>Área</i>					
Urbana	13,1	46,1	61,5	72,8	63,6
Rural	86,9	53,9	38,5	27,3	36,4

<i>Región</i>					
Costa Norte	0,7	0,8	0,8	22,5	15,2
Costa Centro	0,7	4,9	0,4	10,6	8,4
Costa Sur	0,0	2,9	49,0	4,2	5,3
Sierra Norte	0,0	0,9	0,4	8,0	5,6
Sierra Centro	0,0	41,3	2,1	9,7	18,1
Sierra Sur	0,0	31,4	41,0	4,6	13,3
Selva	98,5	10,4	2,5	24,9	21,5
Lima Metropolitana	0,0	7,5	3,8	15,5	12,5
<i>Migración interdistrital</i>					
Extranjero	0,0	0,0	0,0	0,2	0,2
Otro distrito	23,4	43,2	54,4	53,0	49,7
Mismo distrito	76,6	56,8	45,6	46,7	50,2
<i>Migración interprovincial</i>					
Extranjero	0,0	0,0	0,0	0,2	0,2
Otra provincia	15,3	32,5	43,5	33,6	33,3
Misma provincia	84,7	67,5	56,5	66,1	66,5
<i>Migración interdepartamental</i>					
Extranjero	0,0	0,0	0,0	0,2	0,2
Otro departamento	11,0	23,4	37,7	23,9	24,0
Mismo departamento	89,1	76,6	62,3	75,9	75,9
<i>Máximo nivel educativo alcanzado</i>					
Primaria	59,9	43,3	38,1	26,4	32,3
Secundaria	33,6	39,7	42,3	44,8	43,0
Terciaria	6,6	17,0	19,7	28,8	24,7
<i>Nivel educativo del padre</i>					
Sin nivel	54,0	35,0	25,0	15,2	22,8
Primaria incompleta	14,0	27,6	29,4	28,3	27,8
Primaria completa	10,0	12,4	8,7	20,0	16,9
Secundaria incompleta	0,0	1,9	4,4	3,4	2,9
Secundaria completa	4,0	4,0	2,2	9,4	7,3
Superior	2,0	1,9	1,1	5,1	3,9
No sabe	16,0	17,1	29,4	18,5	18,5
<i>Nivel educativo de la madre</i>					
Sin nivel	68,0	62,0	64,1	27,0	40,7
Primaria incompleta	16,0	18,8	16,3	28,0	24,3
Primaria completa	0,0	5,2	4,4	17,8	12,8
Secundaria incompleta	0,0	1,3	0,0	2,9	2,2
Secundaria completa	0,0	1,5	0,0	7,4	5,0
Superior	2,0	0,3	1,1	3,1	2,1
No sabe	14,0	10,9	14,1	13,9	13,0

Nota: la tabla contiene cifras expresadas en porcentaje donde la suma vertical de cada columna y variable es igual a 100%

Fuente: Encuesta Nacional de Hogares 2012. Elaboración propia.

5.5. Análisis multivariado

En esta sección exploramos la relación que existe entre algunas características individuales y contextuales seleccionadas con respecto a las identidades étnicas encontradas a través de las clases latentes. Por la naturaleza de la variable dependiente, una variable con múltiples categorías, proponemos aplicar un modelo basado en una regresión logística multinomial para identificar las características que están asociadas con el uso de determinadas categorías etnorraciales cuando se mantienen constantes otras variables. El grupo que usamos como referencia es el de los individuos identificados como mestizos. Dado que existen diferencias entre las clases latentes según residan en zonas urbanas o rurales, el análisis se hace en cada área geográfica por separado.

Los resultados del modelo multinomial se presentan en la tabla 8. Los coeficientes están expresados en *ratios de riesgo relativo*, o como se les conoce en inglés, «*relative risk ratios*» (Gould, 2000). Este ratio establece la relación entre dos riesgos relativos. Los riesgos relativos se calculan usando los ratios de la probabilidad de que la variable dependiente asuma la categoría de interés en lugar de la categoría de referencia cuando la variable independiente cambia su valor en una unidad. Expresado con notación de fórmula:

$$RRR = \frac{P(y = 1|x+1) / P(y = \text{categoría de referencia} | x+1)}{P(y = 1|x) / P(y = \text{categoría de referencia} | x)}$$

Así, un cambio de una unidad en la variable independiente multiplica el riesgo relativo de la categoría de interés (comparado con la categoría de referencia) por el valor del coeficiente. Cuando el coeficiente asociado al ratio de riesgo relativo asume un valor mayor que 1, esto señala que la variable independiente tiene un impacto positivo sobre la variable dependiente, mientras que, si asume un valor menor que 1, es más probable que la característica asociada tenga un impacto negativo o que opere en la dirección contraria sobre la variable dependiente. Si el coeficiente asume un valor igual o cercano a 1, esto indica la ausencia de relación entre la variable dependiente e independiente.

Al explorar las características demográficas encontramos que en las zonas urbanas los coeficientes asociados al sexo del entrevistado tienen un valor menor que 1. Ello sugiere que las mujeres, en comparación con los hombres, tienen una propensión menor a reportar un patrón de respuestas asociado a alguna de las identidades étnicas indígenas en comparación con una identidad mestiza. Por ejemplo, si calculamos el ratio de la probabilidad de las mujeres urbanas que se identifican con una identidad quechua respecto de la probabilidad de aquellas que escogen una identidad mestiza, y luego repetimos el mismo ejercicio con los hombres, encontramos

que el riesgo relativo es de 0,899. Dado que este valor es menor que 1, vemos que las mujeres en zonas urbanas muestran una menor propensión a exhibir un patrón de respuestas asociado a una identidad quechua. Lo contrario parece ocurrir en las zonas rurales. Sin embargo, como ninguno de los coeficientes alcanza niveles de significancia estadística usualmente aceptados, la relación observada tiene una escasa probabilidad de ser replicada en otra encuesta.

La edad del encuestado agrupada por rangos muestra un patrón distinto entre aquellos que se identifican como quechuas en zonas urbanas respecto de aquellos que lo hacen como mestizos en el mismo ámbito. Observamos que el impacto de la edad fluctúa, aunque no hay un patrón monotónico pero sí estadísticamente significativo; la edad resulta ser más importante para las personas mayores en comparación con los más jóvenes, es decir, aquellos con menos de veinte años. También observamos que la edad es importante para explicar la pertenencia a una identidad amazónica en zonas urbanas, en particular en los tramos de edad más altos, donde los ratios de riesgo relativo alcanzan valores entre 7,8 y 10,8 veces. Sin embargo, el número de casos que representan a los grupos de mayor edad es más escaso entre la población que asume una identidad amazónica, de modo que podría tratarse de un artificio estadístico.

La región natural de residencia, que, como vimos anteriormente, parece estar asociada con ciertas identidades étnicas, tiene un rol importante en el caso de quienes son clasificados dentro de una identidad étnica quechua y aimara. Así, aquellos que residen en zonas urbanas tienen una probabilidad más alta de ser clasificados como quechuas que como mestizos si es que residen en las zonas de la sierra, esto respecto de Lima Metropolitana; dicho impacto también se observa para quienes son clasificados como aimaras. De manera consistente, residir en la región costa o selva respecto de vivir en Lima reduce la probabilidad de que alguien sea clasificado como quechua en comparación con alguien clasificado como mestizo. Dada la concentración de población aimara en la costa sur del país, es consistente que el coeficiente asociado a la región costa resulte significativo dentro de este grupo comparado con los mestizos. Los resultados son similares en la muestra que reside en las zonas rurales, con la salvedad de que el grupo de referencia son las personas que residen en la selva.

La condición migratoria opera de manera esperada: es menos probable que una persona que vive en zonas urbanas y sin experiencia migratoria utilice categorías de respuesta que lo ubiquen en la clase quechua o aimara. Entre la población que vive en zonas rurales ocurre lo contrario: la condición de no migrante se asocia fuertemente con la probabilidad de ser clasificado en una clase amazónica respecto de una clase mestiza, lo que se podría explicar en buena medida como una menor propensión en este grupo poblacional a realizar grandes desplazamientos geográficos.

Tabla 8. Modelo de regresión logística multinomial para identificar las características asociadas a cada clase/identidad latente (reportados como *ratios de riesgo relativo*)

	Muestra urbana			Muestra rural		
	Identidad amazónica	Identidad quechua	Identidad aimara	Identidad amazónica	Identidad quechua	Identidad aimara
	coef/se	coef/se	coef/se	coef/se	coef/se	coef/se
<i>Intercepto/línea de base</i>	0,000 (0,000)	0,414*** (0,082)	0,037*** (0,018)	0,061*** (0,024)	0,280*** (0,053)	0,004*** (0,003)
<i>Género</i>						
Mujer	0,745 (0,363)	0,899 (0,078)	0,815 (0,141)	1,302 (0,289)	0,871 (0,085)	1,172 (0,264)
<i>Grupo de edad (ref.: 14-19)</i>						
20-	4,811 (5,370)	1,658*** (0,259)	1,277 (0,400)	1,298 (0,476)	0,925 (0,154)	0,699 (0,322)
30-	1,183 (1,689)	1,329* (0,218)	1,089 (0,355)	2,046** (0,695)	0,914 (0,150)	1,482 (0,573)
40-	0,000 (0,002)	1,566*** (0,253)	1,003 (0,330)	1,790 (0,657)	0,931 (0,160)	1,885* (0,721)
50-	7,844* (8,811)	1,365* (0,225)	0,914 (0,306)	1,954* (0,791)	1,184 (0,221)	1,787 (0,802)
60-	10,791** (12,528)	1,014 (0,179)	0,675 (0,238)	1,252 (0,580)	1,351* (0,239)	2,879*** (1,163)
<i>Región natural (ref. Lima Metropolitana/Selva)</i>						
Costa	n.e.	0,453*** (0,062)	6,594*** (2,336)	0,000 (0,000)	0,506*** (0,107)	5,326*** (3,027)
Sierra	n.e.	6,118*** (0,731)	6,526*** (2,519)	0,000 (0,000)	7,600*** (1,022)	14,049*** (7,491)
Selva	n.e.	0,598*** (0,091)	0,237* (0,186)			
<i>Condición migratoria</i>						
No es migrante	1,196 (0,598)	0,650*** (0,062)	0,395*** (0,079)	5,882*** (1,541)	0,789* (0,096)	1,070 (0,308)
<i>Nivel educativo (ref. Primaria)</i>						
Secundaria	1,667 (1,105)	0,409*** (0,051)	0,423*** (0,101)	0,699 (0,183)	1,117 (0,137)	1,947** (0,547)
Terciaria	0,958 (0,744)	0,274*** (0,037)	0,316*** (0,082)	0,596 (0,316)	0,800 (0,163)	1,724 (0,749)

Nota: Niveles de significancia: *** p < 0,01, ** p < 0,05, * p < 0,1. El texto «n.e.» indica que no hay suficientes observaciones para estimar el coeficiente (error estándar). El dominio Lima Metropolitana solo incluye áreas urbanas, por lo que la categoría de referencia para el análisis de la muestra rural es la región de la selva.

Fuente: Encuesta Nacional de Hogares 2012. Elaboración propia.

6. MECANISMOS

6.1. El rol del contexto

Muchas de las teorías sobre procesos de autoidentificación enfatizan que las identidades están vinculadas con el contexto en el que operan (Tajfel, 1981). De acuerdo con dichas teorías, existe una relación entre la identidad asumida y el nivel de concentración de la población que muestra cierta característica. Considerando que en el Perú existen grandes espacios y territorios que pueden ser caracterizados como «regiones étnicas indígenas» (Figuroa y Barrón, 2005), es posible esperar que los procesos de autoidentificación estén fuertemente condicionados por la distribución de ciertas características contextuales en las áreas donde residen los individuos. En áreas con alta concentración de poblaciones coétnicas, los individuos más probablemente escogerán una identidad similar a la de la mayoría. Las marcadas diferencias entre las áreas urbanas y rurales nos llevan a considerar que algunos de los análisis se hagan por separado.

En la tabla 9 presentamos la relación entre las características contextuales y los indicadores de autoidentificación étnica. Un primer aspecto a notar son las claras diferencias entre las zonas urbanas y rurales. Así, una persona que indica que habla quechua o aimara y vive en áreas urbanas es probable que viva en un distrito donde, en promedio, entre el 26% y 31% de la población habla una lengua nativa, mientras que en distritos ubicados en zonas rurales ese porcentaje aumenta hasta 68% y 75%, respectivamente. De manera similar, quienes se identifican como quechuas y aimaras según sus antepasados y residen en zonas rurales, tienen una mayor probabilidad de estar expuestos a una mayor concentración de población que habla una lengua nativa. Incluso aquellos que reportan pertenecer a un pueblo indígena en áreas urbanas muestran tasas de exposición a una población de habla nativa notablemente distintas respecto de quienes viven en zonas rurales.

Si exploramos las diferencias según el número de comunidades nativas donde residen los encuestados, los resultados muestran una ligera variación: la proporción de hablantes de otras lenguas nativas, o que se identifican como nativos de la Amazonía, incluso la pertenencia a un pueblo indígena, son mucho mayores en distritos de las zonas donde el promedio de comunidades es marcadamente más alto respecto de la población que reside en zonas urbanas. Así, es más probable que una persona que tiene una lengua nativa y que vive en un distrito de una zona rural se identifique como nativo de la Amazonía.

El tiempo transcurrido desde la migración inicial o la última migración es una variable importante que no es considerada en el análisis. Es posible esperar que a mayor tiempo en el lugar de residencia actual se incremente la probabilidad de que los individuos escojan una identidad étnica-racial que predomine en la comunidad

de residencia y que progresivamente abandonen su etnicidad de origen, en especial en áreas urbanas, que suelen ser el destino final de muchos de los movimientos migratorios.

Tabla 9. Características contextuales y autoidentificación étnica

	% población que habla lengua nativa en el distrito 1/		Total de comunidades nativas en el distrito 2/	
	Área urbana	Área rural	Área urbana	Árearural
<i>Idioma o lengua materna que aprendió en su niñez</i>				
Quechua	26	68	0	1
Aimara	31	75	0	0
Otra lengua nativa	6	57	4	45
Castellano	7	10	1	2
<i>Por sus antepasados y de acuerdo a sus costumbres, usted se considera</i>				
Quechua	22	63	0	1
Aimara	28	72	0	0
Nativo de la Amazonía	5	41	2	33
Mestizo	6	9	1	2
<i>¿Pertenece o se considera de un pueblo indígena?</i>				
Sí	12	44	0	7
No	8	31	1	2

Notas:

1/ Porcentaje de la población de tres años a más cuya lengua materna es quechua, aimara, asháninka u otra lengua nativa dentro del distrito donde reside el encuestada. Fuente: Censo de Población y Vivienda 2007

2/ Número de comunidades indígenas por distrito. Fuente: Censo de Comunidades Nativas 2007.

Fuente: Encuesta Nacional de Hogares 2012. Elaboración propia.

6.2. La transmisión intergeneracional de las identidades etnoraciales

La falta de concordancia entre las preguntas de autoidentificación sugiere que lo que podría estar ocurriendo es que las personas cruzan con mayor fluidez las fronteras etnoraciales. Así, si lo que apreciamos es una ausencia de concordancia en el uso de categorías asociadas con una identidad indígena entre una población que presumimos tiene origen indígena o, dicho de otro modo, si es que hay una menor proporción de indígenas optando por usar categorías relacionadas a una identidad indígena, podemos proponer un abandono progresivo de las categorías etnoraciales relacionadas con una identidad indígena. Si eso fuera cierto, entonces uno esperaría que dicho proceso se inicie en el hogar y sea en cierto modo iniciado por

los padres; así, los padres con antecedentes indígenas mostrarían mayor propensión a reportar una identidad distinta que la de sus hijos, y los hijos mostrarían una tendencia a rechazar una identidad similar a la de sus padres.

Para probar dicha hipótesis se requiere contar con información sobre la identidad racial de los padres de cada uno de los encuestados. Si bien esta es una limitación, notamos que la base de datos sí permite obtener dicha información para la cohorte más joven de individuos, pues una gran mayoría de ellos aún sigue viviendo con sus padres. Explotaremos que la norma social que predomina aún en las familias peruanas es que los hijos tienden a permanecer en la vivienda familiar incluso después de que alcanzan la mayoría de edad, en muchos casos hasta haber asegurado su independencia económica o haber iniciado una nueva unidad familiar. Para ello, limitaremos nuestro análisis a los hijos que tienen entre catorce y veinticinco años de edad y retendremos algunas de sus características etnoraciales, y luego uniremos dichas características con las de sus padres⁶. En la medida en que estamos interesados en la «heredabilidad» de las características etnoraciales, reportamos la probabilidad de que un hijo reporte la misma identidad que su padre o madre.

Así, en la primera línea de la tabla 10 figura el porcentaje de hijos/hijas que respondieron habiéndose identificado como quechuas y tienen padres que también se reportaron como tales: el 79,7% de la generación más joven tiene padres que se reportan como quechuas, mientras que en el caso de las madres el porcentaje es de 83,9%, siendo mucho más altos los porcentajes en zonas rurales que urbanas. Para el caso de los jóvenes autoidentificados como aimaras, los porcentajes son consistentemente altos. En el caso de los jóvenes que se autoidentifican como nativos de la Amazonía, los porcentajes difieren drásticamente. Así, aproximadamente 40% de estos jóvenes usan una categoría de identificación igual a la de sus padres. Ello sugiere un abandono progresivo, en particular entre estos jóvenes en zonas urbanas.

Las tablas 11 y 12 reportan de manera similar el porcentaje de consistencia usando el idioma y la pertenencia a un pueblo indígena. En el caso del idioma, la concordancia entre el idioma reportado por el padre y el reportado por el hijo refleja un caso de transmisión intergeneracional casi perfecta, con porcentajes cercanos al 100%: es a través de los padres que los hijos adquieren el dominio de la lengua predominante en el hogar. En el caso de aquellos jóvenes que se identifican con la categoría de pueblo indígena, 41%-38% de ellos tienen padres/madres que se identifican como tales.

⁶ Es importante notar que la ENAHO no vincula directamente a las personas identificadas como «hijos» con sus respectivos padres o madres biológicos o no. Sin embargo, sí es posible identificar a las personas que dentro del hogar son identificadas como jefe del hogar o esposa, además de su género. Con ello es posible reconstruir de manera bastante precisa las relaciones de filiación entre padres e hijos (Schwartzman, 2007).

Tabla 10. Porcentaje de hijos/hijas clasificados en la misma categoría etnorracial de los padres

	Padre			Madre		
	Total	Urbana	Rural	Total	Urbana	Rural
Quechua	79,7	69,9	89,5	83,9	76,5	92,0
Aimara	78,2	70,1	89,4	81,1	72,2	93,9
Nativo de la Amazonía	42,9	21,4	68,7	43,2	23,5	72,5
Negro/Mulato/Zambo/ Afrodescendiente	18,9	17,2	26,6	22,3	20,9	30,2
Blanco	20,5	22,4	15,8	27,7	26,8	30,2
Mestizo	69,4	70,3	65,0	68,2	69,1	63,5
Otro	40,8	31,6	56,1	42,3	33,3	58,0

Fuente: Encuesta Nacional de Hogares 2012. Elaboración propia.

Tabla 11. Porcentaje de hijos/hijas que reportan el mismo idioma de los padres

	Padre			Madre		
	Total	Urbana	Rural	Total	Urbana	Rural
Quechua	98,5	97,0	98,9	98,6	97,4	98,9
Aimara	96,0	94,7	96,5	96,1	89,8	98,9
Otra lengua nativa	96,2	58,8	97,3	96,8	78,4	97,4
Castellano	78,1	79,4	73,5	79,7	81,0	74,3
Otras lenguas extranjeras	1,4	1,7	0,0	5,6	1,0	42,6

Fuente: Encuesta Nacional de Hogares 2012. Elaboración propia.

Tabla 12. Porcentaje de hijos/hijas que reportan pertenecer a un pueblo al igual que los padres

	Padre			Madre		
	Total	Urbana	Rural	Total	Urbana	Rural
Sí	40,7	37,5	47,4	37,9	35,9	42,6
No	78,9	77,9	81,7	81,8	80,9	84,6

Fuente: Encuesta Nacional de Hogares 2012. Elaboración propia.

7. CONCLUSIONES

Los resultados del análisis de patrones de respuesta, usando análisis de clases latentes sobre las preguntas de autorreconocimiento con ciertos antepasados y costumbres, así como basados en la lengua y pueblo indígena, permiten identificar hasta cuatro posibles identidades etnoraciales. Asociadas a cada una de las identidades, encontramos que la identidad mestiza constituye el grupo de mayor tamaño, con aproximadamente 65% del total de personas de catorce años o más que residen en zonas urbanas y rurales del país. La segunda identidad en importancia, según el tamaño que representa, es la identidad quechua, que alcanza aproximadamente al 29% de la población. Las identidades aimara y amazónica representan aproximadamente el 3,3% y 2,0% de la población, respectivamente. Cada una de estas identidades tiende a mostrar patrones excluyentes; es decir, en contraste con la noción de una multiplicidad de identidades, lo que encontramos son identidades claramente delimitadas a partir de ciertos marcadores.

La composición en términos demográficos de cada una de las identidades es bastante homogénea, tanto por sexo como por edad —aunque existe un ligero predominio de población de las cohortes más antiguas entre los quechuas y aimaras—. Donde se observan diferencias interesantes es en los aspectos relacionadas con la ubicación y movilidad espacial de quienes conforman cada identidad, encontrándose que quienes se adscriben a una identidad amazónica son más propensos a mostrar menor movilidad y mayor concentración en un área geográfica, en contraste con los quechuas y los aimaras, que muestran mayor incidencia de movilidad expresada a través de sus patrones migratorios.

La pregunta sobre los antepasados/costumbres de una persona resulta ser la de mayor utilidad en términos de su capacidad para visibilizar grupos etnoraciales. Además, entre la submuestra más joven, es la pregunta que muestra una mayor «resistencia» en su capacidad de transmisión intergeneracional.

Respecto de la pregunta sobre pueblos indígenas, encontramos que esta no contribuye sustantivamente a la visibilización de las poblaciones de origen quechua y aimara. Si bien ofrecen una clara ventaja en términos de identificar con claridad a las poblaciones con origen en la Amazonía, es escasa su contribución a la identificación de otras poblaciones. Aunque podemos argumentar que el uso de esta pregunta puede responder a una reactividad ante los procesos de exclusión social, no encontramos evidencia clara de que ese sea el caso.

Así, con las herramientas metodológicas usadas, y con la actual modalidad de recolección de datos, no se encuentra evidencia que indique que las identidades étnicas son múltiples y fluidas. Sin embargo, a pesar de la coherencia que se observa dados los patrones de respuestas, las identidades étnicas muestran cierta

heterogeneidad en términos de las dimensiones que las componen. Así, entre quienes son clasificados dentro de identidades andinas, sean estas quechuas o aimaras, el idioma materno y el reconocimiento con los antepasados son condición suficiente para construir dicha identidad, en tanto que en el caso de la identidad amazónica, además de estas dimensiones, la noción de pertenencia a un mismo pueblo indígena juega un rol importante.

En este punto es posible suponer que un ajuste en las herramientas de recolección de datos podría tener consecuencias en los resultados de visibilización, pues podría incrementar la probabilidad de encontrar grupos que, aunque de menor tamaño, reflejen parcialmente la idea de identidades múltiples.

Más aún, si bien el diseño del cuestionario, las preguntas y las categorías de respuesta en su actual formato contribuyen a visibilizar a la población de origen indígena, es poco lo que se puede lograr con las actuales preguntas para delimitar claramente a la población negra/de origen afroperuano. El análisis realizado por Benavides, Sarmiento, Valdivia y Moreno (2013) a una encuesta a la población afrodescendiente indica la potencialidad que el cambio en el diseño de las preguntas tiene para caracterizar mejor a estas poblaciones. Consideramos que aquí se requiere abordar el tema desde dos frentes. Por un lado, es necesario considerar la incorporación de opciones que expandan el abanico de categorías relacionadas con la autoidentificación como afrodescendiente. Por otro, se debe discutir la utilidad de incluir preguntas que recogen algunos rasgos fenotípicos de la población. Esta dimensión ha sido negada en muchas de las investigaciones, aparentemente, en un intento de evitar caer en una reificación biologicista de la noción de raza. Sin embargo, y desde una perspectiva novedosa, Edward Telles y sus colegas (Flores y Telles, 2012; Telles, Urrea y Flores, 2011) han recuperado la importancia de introducir preguntas para capturar las diferencias en el color de la piel y mejorar los procesos de autoidentificación étnica y racial.

Escapa a los alcances de esta investigación el rol de la comprensión e interpretación de cada una de las preguntas sobre autoidentificación. Si, como señala Valdivia (2012), el significado y el sentido de las preguntas no son adecuadamente comprendidos por los encuestados, entonces las respuestas pueden estar midiendo el fenómeno inadecuadamente.

REFERENCIAS BIBLIOGRÁFICAS

- Balarín, M. (2014). Ciudadanía y derechos indígenas. Políticas de Estado para reconocer la diversidad étnica y cultural. En *Inclusión social: diálogos entre la investigación y las políticas públicas*. Lima: GRADE.
- Benavides, M., Sarmiento, P., Valdivia, N. y Moreno, M. (2013). *¡Aquí estamos! Niñas, niños y adolescentes afroperuanos*. Lima: CEDET, Plan Internacional, Unicef.
- Del Popolo, F. (2008). Los pueblos indígenas y afrodescendientes en las fuentes de datos: experiencias en América Latina. Serie Documentos de Proyectos, N.º 197. Cepal.
- Del Popolo, F., Oyarce, A. M., Schkolnik, S. y Velasco, F. (2009). Censos 2010 y la inclusión del enfoque étnico: hacia una construcción participativa con pueblos indígenas y afrodescendientes de América Latina. En *Seminarios y conferencias*. Santiago de Chile: CEPAL, Centro Latinoamericano y Caribeño de Demografía (CELADE), División de Población. Disponible en <http://www.eclac.cl/publicaciones/xml/3/37453/S57SyC-L3095e-P.pdf>
- Figueroa, A. y Barrón, M. (2005). *Inequality, Ethnicity and Social Disorder in Peru*. Working Paper N° 8. Oxford: Crise.
- Flores, R. y Telles, E. (2012). Social stratification in Mexico: disentangling color, ethnicity, and class. *American Sociological Review*, 77(3), 486-494.
- Instituto Nacional de Estadística e Informática (2008). Censos nacionales 2007, XI de población y VI de vivienda: perfil. Segunda edición. Lima: Dirección Técnica de Demografía y Estudios Sociales y Centro de Investigación y Desarrollo del Instituto Nacional de Estadística e Informática. Disponible en <http://censos.inei.gob.pe/Anexos/Libro.pdf>
- Lanza, S. T., Collins, L. M., Lemmon, D. R. y Schafer, J. L. (2007). PROC LCA: A SAS Procedure for Latent Class Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 671-694.
- Llorens, J. A. (2002). Etnicidad y censos: los conceptos básicos y sus aplicaciones. *Boletín del Instituto Francés de Estudios Andinos*, 31(3), 655-680.
- Moreno, M. y Oropesa, R. S. (2012). Ethno-racial Identification in Urban Peru. *Ethnic and Racial Studies*, 35(7), pp. 1220-1247.
- Muthén, B. (2001). Latent variable mixture modeling. En G. A. Marcoulides y R. E. Schumacker (eds.), *New developments and techniques in structural equation modeling*. Mahwah, NJ: Erlbaum.
- Planas, M. E., Cruzado, V., Middelkoop, B., Cárcamo, C. y Richters, A. (2013). Navigating ethnicity-race in Peru: an analytic rationale for measuring multiple self-identification among Indigenous Quechua women. Documento no publicado.
- Schkolnik, S. (2009). *La inclusión del enfoque étnico en los censos de población de América Latina*. Notas de población - Cepal 89, pp. 57-100.
- Schwartzman, L. F. (2007). Does Money Whiten? Intergenerational Changes in Racial Classification in Brazil. *American Sociological Review*, 72(6), 940-963.

- Gould, W. (2000). Interpreting logistic regression in all its forms. *Stata Technical Bulletin*, 53, 19-29.
- Tajfel, H. (1981). *Human Groups and Social Categories: Studies in Social Psychology*. Cambridge: Cambridge University Press.
- Telles, E., Urrea, F. y Flores, R. (2011). *Skin color, racial identity and socioeconomic outcomes in Latin America's pigmentocracy*. Disponible en <http://paa2011.princeton.edu/download.aspx?submissionId=112758>
- Valdivia, N. (2011). *El uso de categorías étnico/raciales en censos y encuestas en el Perú: balance y aportes para una discusión*. Lima: GRADE. Disponible en <http://www.grade.org.pe/publicaciones/detalle/1025>
- Valdivia, N. (2012). Los censos y las encuestas en el Perú y la «visibilización estadística» de la población afroperuana. Disponible en <http://www.grade.org.pe/download/presentaciones/AFROValdivia110512.ppt>
- Wimmer, A. (2007). *How (not) to think about ethnicity in immigrant societies: a boundary making perspective*. Working Paper Series 07-44. Oxford: University of Oxford.

Anexo 1. Transcripción literal de las preguntas usadas para recolectar la autoidentificación étnico-racial

Pregunta sobre idioma o lengua materna

300A. ¿CUÁL ES EL IDIOMA O LENGUA MATERNA QUE APRENDIÓ EN SU NIÑEZ:

- Quechua? 1
Aymara? 2
Otra lengua nativa? _____ 3
(Especifique)
Castellano? 4
Inglés? 5
Portugués? 6
Otra lengua extranjera? _____ 7
(Especifique)
Es sordo mudo 8

Pregunta sobre antepasados/cultura

558C. POR SUS ANTEPASADOS Y DE ACUERDO A SUS COSTUMBRES, ¿UD. SE CONSIDERA:

- Quechua?1
Aymara?2
Nativo o Indígena de la Amazonia?3
Negro/ Mulato/Zambo/Afro peruano?4
Blanco?5
Mestizo?6
Otro? _____ 7
(Especifique)
No Sabe8

Pregunta sobre pertenencia a pueblos indígenas

558D. ¿UD. PERTENECE O SE CONSIDERA PARTE DE UN PUEBLO INDÍGENA?

Sí 1 → 558D1. ¿A QUÉ PUEBLO INDÍGENA PERTENECE?

(Especifique)

- No 2
No sabe 3

Anexo 2. Estadísticas de los modelos de clases latentes

	Número de clases latentes			
	2	3	4	5
Grados de libertad	34	23	12	1
Entropía, R2	0,91	0,94	0,94	0,69
Entropía, sin ajuste	421,30	450,28	514,13	3 273,93
BIC ajustado	1 854,95	1 057,06	265,51	324,21
CAIC	1 942,68	1 190,74	445,15	549,81
BIC	1 921,68	1 158,74	402,15	495,81
AIC	1 778,79	941,01	109,57	128,38
G cuadrado	1 736,79	877,01	23,57	20,38
Loglikelihood	-14 252,78	-13 822,89	-13 396,17	-13 394,57