

**“Focalización Geográfica del Gasto Social: Mapas de Pobreza”**

Javier Escobal \*  
Máximo Torero \*  
Carmen Ponce \*\*

**RED CIES DE POBREZA**

**GRADE-APOYO**

**INFORME FINAL**

**Junio, 2001**

---

\* Investigador Principal, GRADE  
\*\* Investigadora Asistente, GRADE

## INDICE

<b>1. INTRODUCCIÓN</b> .....	<b>3</b>
<b>2. LOS MAPAS DE POBREZA COMO INSTRUMENTOS DE FOCALIZACIÓN DE LAS ESTRATEGIAS DE COMBATE A LA POBREZA</b> .....	<b>7</b>
2.1. LA FOCALIZACIÓN GEOGRÁFICA EN EL MARCO DE FOCALIZACIÓN DE LAS ESTRATEGIAS DE COMBATE A LA POBREZA .....	7
2.1.1. <i>Los instrumentos de focalización en los programas de combate a la pobreza</i> .....	7
2.1.2. <i>La focalización geográfica: racionalidad, ventajas y limitaciones</i> .....	8
2.2. LOS MAPAS DE POBREZA UTILIZADOS COMO INSTRUMENTOS DE FOCALIZACIÓN GEOGRÁFICA EN EL PERÚ DURANTE DÉCADA DE LOS NOVENTA.....	11
2.2.1 <i>Descripción de los principales mapas de pobreza disponibles</i> .....	12
2.2.2 <i>Avances y Limitaciones de los mapas en el Perú</i> .....	15
<b>3. MARCO METODOLÓGICO PARA LA ELABORACIÓN DE MAPAS DE POBREZA MEDIANTE LA UTILIZACIÓN COMPLEMENTARIA DEL CENSO DE POBLACIÓN Y VIVIENDA Y UNA ENCUESTA DE HOGARES SOBRE CONDICIONES DE VIDA Y POBREZA</b> .....	<b>19</b>
3.1 ESTIMACIÓN DEL NIVEL DE GASTO PER CÁPITA O DE LA PROBABILIDAD DE SER POBRE DEL HOGAR CENSAL .....	20
3.1.1 <i>Definición de la unidad de análisis y compatibilización de las fuentes de información utilizadas (censo y encuesta)</i> .....	20
3.1.2 <i>Estimación de los parámetros que resumen la relación funcional entre las variables socio económicas y demográficas y el nivel de gasto per cápita o la probabilidad de ser pobre de los hogares muestreados en la encuesta</i> .....	24
3.1.3 <i>Interpolación de los parámetros estimados en 3.1.2 sobre la base censal de hogares</i> .....	29
3.2 AGREGACIÓN DE LAS ESTIMACIONES POR HOGAR: TASA DE POBREZA Y ERROR DE PREDICCIÓN A NIVELES DIFERENCIADOS DE AGREGACIÓN GEOGRÁFICA .....	30
3.2.1 <i>Estimación de la Tasa de Pobreza a niveles geográficos determinados</i> .....	32
3.2.2 <i>Cálculo del error estándar de la tasa de pobreza estimada a niveles geográficos determinados</i> .....	33
<b>4 APLICACIÓN PRÁCTICA DE LA METODOLOGÍA DE ESTIMACIÓN DE INDICADORES DE BIENESTAR A NIVELES REDUCIDOS DE AGREGACIÓN GEOGRÁFICA: PRINCIPALES RESULTADOS</b> .....	<b>36</b>
4.1 ESTIMACIÓN DEL NIVEL DE GASTO PER CÁPITA O DE LA PROBABILIDAD DE SER POBRE DEL HOGAR CENSAL .....	36
4.1.1. <i>Definición de la unidad de análisis y compatibilización de las fuentes de información utilizadas (censo y encuesta)</i> .....	37
4.1.1 <i>Estimación de los parámetros que resumen la relación funcional entre las variables socio económicas y demográficas y el nivel de gasto per cápita o la probabilidad de ser pobre de los hogares muestreados en la encuesta</i> .....	38
4.2 AGREGACIÓN DE LAS ESTIMACIONES POR HOGAR: TASA DE POBREZA Y ERROR DE PREDICCIÓN A NIVELES DIFERENCIADOS DE AGREGACIÓN GEOGRÁFICA .....	46
<b>5. CONCLUSIONES Y RECOMENDACIONES METODOLÓGICAS</b> .....	<b>52</b>
<b>6. BIBLIOGRAFÍA</b> .....	<b>59</b>
<b>7. BASES DE DATOS</b> .....	<b>62</b>
<b>ANEXO 1</b> .....	<b>63</b>
<b>ANEXO 2</b> .....	<b>74</b>
<b>ANEXO 3</b> .....	<b>81</b>

## 1 INTRODUCCIÓN

Los mapas de pobreza constituyen en la actualidad uno de los instrumentos más importantes para la focalización del gasto destinado al combate a la pobreza en países en desarrollo, típicamente caracterizados por presentar severos problemas de pobreza y limitaciones en el presupuesto disponible para combatirla. En particular, en el Perú, desde inicios de los noventa, parte importante del gasto enmarcado en las estrategias focalizadas de lucha contra la pobreza ha sido asignado haciendo uso de este tipo de instrumentos<sup>1</sup>. Casos típicos del uso de este tipo de metodología son los mapas de pobreza utilizados por el Fondo de Compensación y Desarrollo Social (FONCODES) y por el Proyecto para la Focalización de la Inversión Económica y Social (PROFINES).

Una de las principales características de los mapas de pobreza, y por la cual se ha incrementado su uso como instrumentos de focalización del gasto público han sido sus bajos costos de elaboración, implementación y monitoreo, especialmente cuando se les compara con aquellos costos exigidos por instrumentos alternativos de focalización (v.g. aquellos asociados a la focalización individual). La mayoría de los proyectos y programas que forman parte de la lucha contra la pobreza tienen algún tipo de estrategia de focalización geográfica, al menos una focalización gruesa asignando un mayor gasto hacia los departamentos más pobres.

Sin embargo, como ha sido señalado en diversos estudios y se reitera en el presente documento, los mapas de pobreza también presentan limitaciones conceptuales y operativas nada desdeñables, que se deben tomar en cuenta para reducir en lo posible los problemas de filtración y subcobertura. Una de las principales preocupaciones, y en las que se centra este estudio, radica en que la ganancia en eficiencia, derivada de la utilización de este instrumento y que se sustentan en la homogeneidad al interior de los agregados geográficos respectivos, se ve restringida por las limitaciones en precisión de las estimaciones de pobreza. Esto se debe a que tal supuesto de homogeneidad es válido solo a niveles reducidos de agregación geográfica, a los cuales la información disponible (y con ella la precisión de las estimaciones) es típicamente insuficiente.

En este sentido, la pregunta que guía la presente investigación es ¿cómo mejorar la precisión de los indicadores de pobreza manteniendo niveles reducidos de agregación geográfica?. Al respecto, este documento pretende ser un aporte metodológico al identificar aspectos que se deberían tomar en cuenta al construir los mapas de pobreza. El trabajo se centra en cuatro aspectos metodológicos que pueden contribuir a mejorar la eficacia de los mapas como instrumento de focalización, (a) la utilización de fuentes de información complementarias a la censal que permitan obtener un indicador directo del nivel de bienestar de los hogares (en base a los niveles de ingreso o consumo, haciendo uso de la metodología de línea de pobreza); (b) la exploración de técnicas econométricas que permitan optimizar el uso de esta información, es decir, que mejoren la precisión de las estimaciones efectuadas incorporando aspectos como los marcos muestrales de las encuestas, formas funcionales e incluyendo información adicional que permita la minimización de errores de filtración y subcobertura; (c) la apropiada interpolación en la base de hogares censales utilizando los parámetros estimados a partir

---

1 Ver Saavedra y Maruyama (1999); Chacaltana (2000).

de las fuentes de información complementarias; y (d) la agregación del indicador de bienestar estimado para el hogar a niveles diferenciados de agregación geográfica y el desarrollo de una metodología de cálculo del error de predicción de la tasa de pobreza correspondiente. De esta manera, se pretende mejorar la precisión del indicador de pobreza estimado (tasa de pobreza) así como conocer los intervalos de confianza del mismo, lo cual debería redundar en una mayor precisión del ordenamiento de pobreza que suele proponerse como parte del proceso de focalización <sup>2</sup>.

Las fuentes de información más detalladas disponibles para el Perú, y que permiten conocer los niveles de gasto de los hogares y con ello estimar si los hogares son o no pobres, son las encuestas de hogares (la Encuesta Nacional de Hogares – ENAHO o la Encuesta Nacional de Hogares sobre Niveles de Vida – ENNIV). Estas encuestas, sin embargo son representativas únicamente a nivel de dominios geográficos extensos (por ejemplo, Costa Urbana, Costa Rural, Sierra Urbana, Sierra Rural, Selva Urbana, Selva Rural y Lima Metropolitana) y no son representativas de departamentos y mucho menos de distritos o centros poblados. Obviamente a este nivel de agregación mayor se observa una alta heterogeneidad en los niveles de consumo de los hogares, y con ello en las tasas de pobreza estimadas al interior de cada dominio, lo que implica que un mapa de pobreza obtenido en base a encuestas de hogares no redundaría en un incremento real de la eficiencia en la asignación del gasto público focalizado, en tanto no es capaz de diferenciar y priorizar entre áreas al interior de estos dominios.

En este contexto, el estudio buscará presentar un método para generar mapas de pobreza a distintos niveles de agregación (departamental, provincial y distrital) combinando información censal con información de encuestas de hogares, siguiendo la metodología desarrollada por Hentschel et al.(2000) y Elbers et al. (2000). Con este fin, y para validar la metodología presentada se realiza un ejercicio, tomando como unidad de análisis el hogar y utilizando dos bases de datos. La primera de ellas se generó a partir de los reportes (individuales y por hogar) recogidos en la Encuesta Nacional de Hogares 1998 del IV Trimestre. Esta base de datos permite estimar la relación existente entre el nivel de gasto per cápita del hogar y las características del mismo (en especial las referidas al acceso a activos privados y públicos). La segunda, por su parte, generada a partir de información del 5% de los hogares encuestados en el Censo IX de Población y IV de Vivienda de 1993, permitió efectuar predicciones de los niveles de gasto de los hogares censales. Estas predicciones se efectuaron interpolando las ecuaciones estimadas en base a la relación entre gasto y las características de los hogares obtenida de la ENAHO. En base a estas predicciones se efectuaron los cálculos de las tasas de pobreza a nivel de centro poblado, distrito, provincia, departamento y dominio geográfico.

Sin embargo, es de esperar que la confiabilidad de las estimaciones se incremente conforme crece el tamaño poblacional de los referidos agregados geográficos<sup>3</sup>. Este documento presenta la metodología de cálculo del error estándar o error de predicción de las estimaciones referidas, (desarrollada, al igual que la metodología de estimación de las tasas de pobreza, por Hentschel, Jean Lanjouw, Peter Lanjouw, Poggi y Elbers),

---

2 En particular, este tipo de metodología permite efectuar ordenamientos geográficos de pobreza incluso entre áreas geográficas con indicadores similares, al contar con criterios adicionales como el proporcionado por los intervalos de confianza.

3 Como se mostrará en el documento, los indicadores de pobreza a nivel de centro poblado no son confiables, debido a los enormes errores de predicción que involucran.

la cual permite ayudar a la definición del nivel de agregación geográfica mínimo requerido para validar las estimaciones de pobreza, así como mejorar la precisión del ordenamiento de agregados geográficos y con ellas los coeficientes de asignación de gasto público que se puedan derivar del mismo. El error estándar referido está integrado por tres componentes, (a) el error idiosincrático, asociado a variables no contempladas en el modelo estimado y que son relevantes para la sub-población analizada; (b) el error del modelo que incorpora a su vez errores derivados del marco muestral utilizado en la encuesta de hogares y de la varianza de los estimadores obtenidos; y (c) el error de cómputo, derivado de la necesidad de, en algunos casos, obtener un cálculo aproximado y no exacto del indicador de pobreza<sup>4</sup>.

Es importante destacar que la propuesta metodológica que se plantea en este documento pretende facilitar una rápida actualización de los mapas de pobreza y de asignación de gasto, tan pronto como se desarrollen nuevas encuestas de hogares, y se tenga nueva información del Censo Nacional X de Población y V de Vivienda que el INEI viene programando. En este sentido, el principal aporte de este trabajo no es construir un nuevo mapa de pobreza, sino aportar en el desarrollo de una estrategia metodológica que permita una actualización continua de este tipo de instrumento de focalización, así como su validación a distintos niveles de agregación geográfica.

Finalmente, es importante establecer el contexto en el que se deben colocar los aportes metodológicos que se presentan en este documento. En primer lugar, en referencia a la estrategia de focalización como parte medular de los programas de combate a la pobreza, si bien se reconoce que consta de varias etapas y es transversal al ciclo de proyectos, es importante destacar que el presente trabajo pretende referirse exclusivamente a la primera etapa consistente en la elaboración del instrumento de focalización geográfica, en este caso, del mapa de pobreza. En ese sentido es importante no confundir la propuesta metodológica que aquí se plantea con el desarrollo de una estrategia de focalización, de la cuál el mapa es sólo un instrumento de focalización de programas de combate a la pobreza. En segundo lugar, es importante anotar que el documento hace referencia a metodologías que se basen en indicadores de insuficiencia de consumo así como de carencias en el acceso a bienes o servicios, en tanto *proxys* de pobreza. De esta manera, se dejará de lado la discusión relacionada a la superioridad o no de mapas de pobreza que se basen en indicadores de carencias que constituyan en sí mismos indicadores de definición de acciones del programa, es decir, de indicadores que tengan un correlato directo con las líneas de acción de los programas particulares.

El documento consta de cinco secciones incluyendo esta introducción. La segunda sección evalúa los mapas de pobreza como instrumento de focalización y se divide a su vez en dos partes. La primera parte presenta una revisión básica de los mapas de pobreza en el marco de los instrumentos de focalización utilizados en la actualidad, haciendo especial énfasis en sus principales cualidades y limitaciones. La segunda parte presenta una revisión crítica de los principales mapas de pobreza que vienen siendo utilizados como instrumentos de focalización del gasto público en el Perú, o que, aunque aun no operativos, se planea sean utilizados pronto.

---

4 Según Elbers, J. Lanjouw y P. Lanjouw, este componente suele ser poco significativo en el error total de predicción (para el caso de Ecuador se estima en 0.001 lo que representa menos del 2% de la tasa media de pobreza). Considerando que su cálculo es especialmente complejo al requerir un proceso de simulaciones, se ha optado por no incorporarlo en el cálculo que se presenta en este documento.

En la tercera sección se presenta la metodología propuesta para la elaboración del mapa de pobreza, así como la metodología de cálculo del error de las estimaciones que se derivan de esta propuesta. En la cuarta sección, se desarrolla una aplicación práctica de la metodología de estimación de indicadores de bienestar a niveles reducidos de agregación geográfica de acuerdo a la metodología desarrollada en la tercera sección. Esta aplicación práctica utiliza la información de la encuesta de hogares ENAHO del cuarto trimestres de 1998 y el 5% del censo IX de Población y vivienda realizado en 1993. Finalmente, en la última sección, se presentan las conclusiones y recomendaciones metodológicas al momento de elaborar los mapas de pobreza. Asimismo, el documento incluye tres anexos metodológicos en donde se señalan los principales ajustes realizados para efectuar los distintos cálculos hechos en la aplicación práctica de esta investigación.

## **2 LOS MAPAS DE POBREZA COMO INSTRUMENTOS DE FOCALIZACIÓN DE LAS ESTRATEGIAS DE COMBATE A LA POBREZA**

### **2.1 La focalización geográfica en el marco de focalización de las estrategias de combate a la pobreza**

La presente sección busca presentar una rápida revisión al tema de elaboración y utilización de mapas de pobreza como instrumentos de focalización geográfica en las estrategias de combate a la pobreza. En primer lugar, se enmarca a la focalización geográfica en la programación del gasto público focalizado. En seguida, se presentan las principales ventajas y limitaciones que ofrecen los mapas de pobreza como instrumentos de focalización. Finalmente, se esbozan algunas de las principales ideas que constituyen el sustento teórico de la utilización de este tipo de instrumento para mejorar la eficacia del gasto público focalizado.

#### **2.1.1 Los instrumentos de focalización en los programas de combate a la pobreza**

Los instrumentos de focalización<sup>5</sup> implementados en la actualidad constituyen un elemento fundamental, tanto en la asignación eficiente del gasto público como en la ejecución misma de dicho gasto, en tanto la limitación de recursos para la atención al problema de la pobreza no suele permitir implementar programas de cobertura universal.

Los instrumentos de focalización utilizados actualmente para la asignación de presupuesto y ejecución de proyectos de combate a la pobreza pueden ser de dos tipos<sup>6</sup>:

- 1. Por indicadores socio-económicos.** Los instrumentos de focalización de este tipo se basan en indicadores socio-económicos que permiten identificar a la población objetivo de los programas, es decir, que ofrecen criterios de discriminación entre situación de pobreza y no pobreza. Estos instrumentos pueden ser de dos tipos, dependiendo del nivel de agregación al que se refieran los indicadores:
  - a. Instrumentos de focalización geográfica.** Tienen como objetivo identificar áreas geográficas en las cuales se concentra la población con problemas de pobreza, estableciendo ordenamientos geográficos que faciliten la priorización en la asignación del gasto público. Como se indica a lo largo del documento, este tipo de focalización es útil a niveles reducidos de agregación geográfica, que se caracterizarían por la relativa homogeneidad de su población.
  - b. Instrumentos de focalización individual o familiar.** Tienen como objetivo identificar individuos u hogares en situación de pobreza, por medio de indicadores elaborados con información individual. Si bien este tipo de instrumento es ideal, en tanto reconoce la heterogeneidad al interior de los

5 Como es conocido, los instrumentos de focalización de los programas públicos tienen como finalidad facilitar la identificación de la población objetivo a la que estos pretenden atender, típicamente la población en situación de pobreza o de imposibilidad de acceso a determinados bienes o servicios indispensables para mantener niveles de vida aceptables.

6 Respecto de la tipología de mecanismos de focalización, ver Francke y Guabloche 1995; Grosh 1994; Lipton y Ravallion 1993; Haddad y Kanbur 1992.

agregados geográficos, generalmente no se dispone de información confiable a niveles de desagregación tan elevados y en los casos en que se dispone esta es específica a pequeñas zonas del país.

**2. Por auto selección.** Este mecanismo de focalización se caracteriza por la ausencia de instrumentos explícitos de selección o exclusión de beneficiarios de los programas. Típicamente, los programas que hacen uso de este mecanismo de focalización cuentan con esquemas de desincentivos a la participación de poblaciones no pobres, por lo que propician la autoexclusión de la población no pobre sin incurrir en costos de administración y monitoreo de focalización. Cabe notar, sin embargo, que la eficacia de esta estrategia de focalización (basada en la eficacia de los esquemas de desincentivos referidos) suele limitarse a pocos proyectos, generalmente orientados a poblaciones en situación de extrema pobreza, involucrando transferencias de bienes o servicios de baja calidad o que son demandados únicamente por la población más necesitada.

En adelante, se hará mención exclusivamente a instrumentos explícitos de focalización por indicadores socio-económicos, es decir, a instrumentos de focalización geográfica e individual, dejando de lado los mecanismos de focalización por auto selección.

Respecto de los tipos de focalización por indicadores socio-económicos, cabe señalar que la elección del instrumento de focalización a utilizar en un programa particular sigue, en general, criterios de costo-efectividad que pueden conducir a la utilización de ambos tipos de instrumento dependiendo de la estrategia utilizada en cada etapa de diseño e implementación del programa. Por un lado, ambos instrumentos de focalización (geográfica e individual) persiguen el mismo objetivo: identificar con la mayor precisión posible a la población en situación de pobreza que debe ser atendida; y por otro, la utilización de ambos tipos de instrumento conlleva errores de filtración (individuos u hogares no pobres son beneficiados directamente con las acciones del proyecto) y subcobertura (individuos u hogares en situación de pobreza son excluidos de los beneficios del proyecto). La elección del tipo de instrumento dependerá de lo costo-efectivo que cada uno sea, es decir, de la precisión con la cual sean capaces de identificar a la población objetivo dado un presupuesto limitado. Así, por un lado, los instrumentos de focalización geográfica (mapas de pobreza) han mostrado ser costo-efectivos a niveles de agregación a los que los costos de elaboración e implementación de instrumentos individuales son prohibitivos. No obstante lo anterior, conforme se reduce el nivel de agregación al que es necesario identificar a la población objetivo, la pérdida de precisión de los instrumentos geográficos respecto de los individuales empieza a revertir las ventajas de un instrumento sobre el otro<sup>7</sup>.

### **2.1.2 La focalización geográfica: racionalidad, ventajas y limitaciones**

La focalización geográfica, es decir la focalización que tiene como objetivo la identificación y atención de áreas pobres, surge por el reconocimiento de diferencias significativas en los niveles de vida observados entre áreas geográficas. Diversos estudios sobre el tema, como el de Ravallion y Woodon (1997) o el de Bigman y Fofack

---

<sup>7</sup> Un ejemplo de este tipo de razonamiento lo da FONCODES, que utiliza un mapa de pobreza para asignar fondos a nivel distrital, utilizando de manera creciente instrumentos más individuales al interior de esos distritos.

(2000), señalan que las diferencias observadas entre los niveles de vida promedio de diferentes comunidades parecerían ser demasiado grandes para ser explicadas por diferencias en características individuales o familiares observables, no geográficas y móviles (variables demográficas, laborales, educacionales, etc.) y que por consiguiente es importante incluir características geográficas. Al respecto, Bigman y Fofack (2000: 130), señalan algunas de las principales características geográficas que podrían explicar las disparidades observadas entre áreas diferentes, más aún estas características pueden estar potenciadas o originadas por la presencia de una geografía natural adversa (ver Escobal y Torero, 2000). Entre las principales características están:

1. La baja calidad de los servicios públicos, particularmente en educación y salud, que impide la acumulación de capital humano y con ello la capacidad de generar ingresos.
2. La deficiente condición de la infraestructura rural, particularmente carreteras y otras vías de comunicación, que limita el intercambio y retarda la inversión local (Biswanger, Khandker, y Rosenzweig 1993).
3. El bajo nivel de capital social en comunidades pobres que retarda la difusión y adopción de nuevas tecnologías agropecuarias, reduciendo así la capacidad de generación de ingresos (Foster y Rosenzweig 1995).
4. La distancia de centros urbanos que inhibe el comercio, la especialización en la producción, y el acceso a crédito que normalmente se concentra en las capitales de ciudades.
5. Los obstáculos a la migración hacia áreas más desarrolladas, resultado de factores económicos, demográficos y culturales diversos. Al respecto, un problema importante es el de la escasa información disponible que facilite la toma de decisión (migrar o no migrar, cuándo, hacia dónde y cómo hacerlo).

Adicionalmente a la importancia que este tipo de variables puedan tener en la explicación de las diferencias entre niveles de vida de diferentes áreas geográficas, Ravallion y Woodon (1997), para el caso de Bangladesh y Escobal y Torero (2000), para el caso del Perú, señalan que existe una segunda fuente potencial de disparidades entre niveles de vida de hogares similares residentes en distintos agregados geográficos. Esta segunda fuente es un conjunto de variables no geográficas que no son fácilmente observables y que, estando correlacionadas geográficamente, puedan estar condicionando las diferencias observadas entre áreas de residencia distintas. Así pues el proceso de identificación de poblaciones pobres debe tener en cuenta no sólo variables geográficas y variables individuales no geográficas y móviles observables, sino también a éstas variables que podrían ser incorporadas al análisis con métodos de econometría espacial (ver Escobal y Torero, 2000).

Las ideas expuestas constituyen parte esencial del sustento conceptual de las estrategias de focalización geográfica en los programas de combate a la pobreza, y en consecuencia de la elaboración de mapas de pobreza y del diseño de líneas de acción de tales programas<sup>8</sup>. Cabe enfatizar, sin embargo, que si bien los mapas de pobreza constituyen el instrumento por excelencia de las estrategias de focalización geográfica, estos pueden ser utilizados como instrumentos de focalización de estrategias de focalización no geográficas, es decir, estrategias que identifiquen líneas de acción que atiendan deficiencias individuales de los hogares pobres. En este contexto, usualmente los mapas

---

8 Una explicación teórica más exhaustiva se puede encontrar en Ravallion y Woodon (1997).

de pobreza constituyen el primer instrumento utilizado en la estrategia de focalización, siendo complementados posteriormente con instrumentos de focalización individual.

Como se mencionó anteriormente, la ventaja más obvia de la utilización de instrumentos de focalización geográfica radica en sus bajos costos de implementación y monitoreo, en relación con otros instrumentos de focalización. Adicionalmente, sin embargo, es posible enumerar una serie de virtudes igualmente importantes. Bigman y Fofack (2000: 134) detallan diversas ventajas de la utilización de instrumentos de focalización geográfica, entre las cuales resaltan las siguientes:

- Provee criterios claros de identificación de la población objetivo y elude problemas derivados de restricciones en la información disponible que impiden la utilización de otros instrumentos de focalización.
- Al ser relativamente fácil de monitorear y administrar, facilita la participación activa de instituciones locales y organizaciones no gubernamentales.
- Su influencia en el comportamiento de los hogares potencialmente beneficiarios es relativamente pequeña, en tanto es difícil y costoso para un hogar cambiar su lugar de residencia con el objeto de participar de los beneficios de un proyecto focalizado geográficamente.
- Es posible mejorar la focalización producto de la utilización de este instrumento, combinando el criterio geográfico con otros criterios de elegibilidad basados en características individuales o familiares.

No obstante el reconocimiento de las virtudes de la focalización geográfica en la eficiencia de la asignación del gasto público, es fundamental reconocer que estas virtudes no se hacen efectivas a cualquier nivel de agregación geográfica. Al respecto, investigadores abocados al tema señalan en diversos estudios que las ganancias derivadas de este tipo de focalización son significativas únicamente cuando es utilizada a niveles de agregación bastante reducidos, a los cuales la población es relativamente homogénea<sup>9</sup> (Baker y Grosh, 1994; Ravallion y Woodon, 1997; Ravallion y Lipton, 1993). En particular, algunas de las principales limitaciones de los mapas de pobreza son<sup>10</sup>:

- La escasa información disponible a nivel de hogares que permita elaborar indicadores a niveles de agregación suficientemente detallados, en especial la información que permita conocer niveles de consumo de los hogares que sirvan de base para la construcción de indicadores directos de bienestar (como los derivados de metodologías de línea de pobreza; incidencia, brecha, severidad).
- La metodología de elaboración de indicadores de pobreza basados en estimaciones de niveles de consumo de los hogares (comúnmente efectuadas mediante encuestas de hogares, representativas a niveles de agregación bastante elevados), induce a problemas de filtración y subcobertura a diferencia de programas de focalización individual.
- En el caso de mapas de pobreza elaborados en base a indicadores complejos (que combinan diversas variables) se presenta normalmente problemas de selección arbitraria de ponderaciones en base a las cuales estos indicadores son calculados.

---

<sup>9</sup> En el caso peruano, por ejemplo, el nivel de agregación al que son utilizados los mapas de pobreza es típicamente el distrital.

<sup>10</sup> Posteriormente se presenta un análisis más detallado de las limitaciones de este tipo de instrumentos en el marco de la experiencia peruana.

- La utilidad de los mapas a lo largo del tiempo requiere del desarrollo de sistemas de actualización permanente que incluyen: (a) actualización de las bases censales; (b) la actualización de las relaciones funcionales en el caso de efectuar interpolaciones (como se propone en el presente documento) ; y (c) metodologías que permitan complementar las bases censales con información actualizada, no necesariamente censal.
- La precisión del ordenamiento de áreas geográficas de un mapa de pobreza depende de la homogeneidad de la población al interior de los mismos, por lo que su utilidad está limitada por el nivel de agregación geográfico para el que sea utilizado. Esta limitación puede ser manejada haciendo uso de instrumentos complementarios de focalización (como los de focalización individual).

## **2.2 Los mapas de pobreza utilizados como instrumentos de focalización geográfica en el Perú durante década de los noventa**

El primer antecedente de los mapas de pobreza construidos con el objetivo de servir como instrumento de focalización del gasto público, es el mapa de pobreza elaborado por el Banco Central de Reserva del Perú (BCRP) el año 1981 en base al Censo de 1972 y otras fuentes estadísticas como la Encuesta Nacional de Consumo Alimentario (ENCA). Este mapa fue elaborado en el marco del programa de distribución de Bonos de Alimentación. Si bien no llegó a ser utilizado, sirvió como primer antecedente y motivó la elaboración del mapa de pobreza de 1986, también a cargo del BCRP, el cual utilizó como base el Censo de 1981 (FONCODES, 2000: 4)<sup>11 12</sup>.

No obstante lo anterior, los mapas de pobreza adquieren relevancia especial como instrumento de asignación geográfica del gasto público recién el año 1991 con la creación de FONCODES, primer programa orientado específicamente a la lucha contra la pobreza. Como indican Saavedra y Maruyama (1999: 20) éste fue el primer organismo estatal en elaborar y utilizar un mapa de pobreza como instrumento oficial de focalización presupuestal. Actualmente, se cuenta con varios mapas de pobreza elaborados con información del Censo IX de Población y IV de Vivienda del año 1993 (en algunos casos complementados con otras fuentes como el Censo de Talla de 1993, encuestas del Ministerio de Salud, estadísticas de transportes y comunicaciones del año 1996).

En esta sección se presenta una breve revisión crítica de los principales mapas de pobreza que están siendo utilizados (o en su defecto que se planea sean utilizados pronto), en la asignación de gasto público focalizado de combate a la pobreza. Esta revisión consta de dos partes, una primera parte descriptiva que señala los indicadores en base a los cuales han sido elaborados los referidos mapas; y una segunda parte que presenta algunas observaciones sobre avances y limitaciones de estos mapas de pobreza en el marco de las que fueron señaladas en la sección anterior.

11 Cabe mencionar, que en el Perú se cuenta con mapas de pobreza desde el año 1977; sin embargo, estos mapas habrían sido elaborados con objetivos de descripción y análisis de la pobreza, antes que como instrumentos de focalización de gasto público. Dos de los mapas elaborados con anterioridad a 1981 y que tuvieron amplia difusión fueron los elaborados por Richard Webb y Carlos Amat y León en base al Censo de 1972.

12 Posteriormente a la elaboración de este mapa, el Instituto Nacional de Estadística e Informática (1985-1986) y el Instituto Cuánto (1988) elaboraron nuevos mapas de pobreza.

### 2.2.1 Descripción de los principales mapas de pobreza disponibles

Respecto de los mapas de pobreza elaborados en base al Censo de 1993, cabe resaltar dos grupos: el grupo de los mapas que han estado siendo utilizados por instituciones gubernamentales en la asignación presupuestal (y que han experimentado cambios importantes a lo largo de la década); y el grupo de los mapas que, aunque no operativos, han sido elaborados para orientar en el futuro la asignación presupuestal del gasto público en los programas de lucha contra la pobreza.

#### *Mapas operativos*

A continuación se presenta una breve enumeración de los principales mapas de pobreza que son utilizados actualmente por el sector público en la asignación presupuestal para el combate de la pobreza (ver la Tabla 1).

1. Mapa de Pobreza del Instituto Nacional de Estadística e Informática (INEI)  
Este mapa utiliza 5 indicadores de Necesidades Básicas Insatisfechas (NBI):
  - Porcentaje de hogares con inasistencia escolar
  - Porcentaje de viviendas inadecuadas
  - Porcentaje de hogares con hacinamiento
  - Porcentaje de viviendas sin servicio de desagüe
  - Porcentaje de hogares con alta dependencia económica

La fuente de información en la que se basa es el Censo IX de Población y IV de Vivienda del año 1993.

2. Mapa de Pobreza de Fondo Nacional de Compensación y Desarrollo Social – FONCODES – 1996  
Este mapa construye un índice sintético de pobreza para cada distrito (1,793); clasifica los distritos en 5 estratos (pobre extremo, muy pobre, pobre, regular y aceptable).  
Los indicadores utilizados son:
  - Tasa de analfabetismo
  - Porcentaje de hogares con inasistencia escolar
  - Porcentaje de viviendas inadecuadas
  - Porcentaje de hogares con hacinamiento
  - Porcentaje de viviendas sin servicio de agua potable
  - Porcentaje de viviendas sin servicio de desagüe
  - Porcentaje de viviendas sin electricidad
  - Tasa de desnutrición crónica
  - Porcentaje de hogares con alta dependencia económicaLas fuentes de información en las que se basa son el Censo IX de Población y IV de Vivienda del año 1993 y el Censo de Talla del año 1993.

**Tabla 1**  
**Mapas de Pobreza Desarrollados en el Perú**

<b>MAPA</b>	<b>Mapa de Pobreza del Instituto Nacional de Estadística e Informática (INEI)</b>	<b>Mapa de Pobreza de Fondo Nacional de Compensación y Desarrollo Social – FONCODES – 1996</b>	<b>Mapa de la Estrategia Focalizada de Lucha contra la Pobreza Extrema del Ministerio de la Presidencia (PROFINES)</b>	<b>Mapa del Ministerio de Economía y Finanzas</b>	<b>Mapa del FONCODES - 2000</b>
<b>PRODUCTO</b>	Mapa de Necesidades Básicas Insatisfechas	Índice sintético de pobreza para cada distrito	Ordenamiento a nivel distrital, y por zonas censales y centros poblados	Ranking distrital	Índice sintético de pobreza para cada distrito
<b>INDICADORES UTILIZADOS</b>	<ul style="list-style-type: none"> <li>▪ Inasistencia escolar</li> <li>▪ Viviendas inadecuadas</li> <li>▪ Hacinamiento en las viviendas</li> <li>▪ Viviendas sin servicio de desagüe</li> <li>▪ Alta dependencia económica</li> </ul>	<ul style="list-style-type: none"> <li>▪ Analfabetismo</li> <li>▪ Inasistencia escolar</li> <li>▪ Viviendas inadecuadas</li> <li>▪ Hacinamiento en las viviendas</li> <li>▪ Viviendas sin servicio de agua potable</li> <li>▪ Viviendas sin servicio de desagüe</li> <li>▪ Viviendas sin electricidad</li> <li>▪ Desnutrición crónica</li> <li>▪ Alta dependencia económica</li> </ul>	<ul style="list-style-type: none"> <li>▪ Primera etapa: ordenamiento a nivel distrital sobre la base del indicador de FONCODES</li> <li>▪ Segunda etapa: ordenamiento por zonas censales y centros poblados sobre la base del mapa de Necesidades Básicas Insatisfechas del INEI</li> </ul>	<ul style="list-style-type: none"> <li>▪ Déficit de consumo de la población pobre</li> <li>▪ Desigualdad en la distribución del gasto</li> <li>▪ Severidad de la pobreza</li> <li>▪ Inversión en cada distrito durante el período 1993-1999 en programas de lucha contra la pobreza.</li> </ul>	<ul style="list-style-type: none"> <li>▪ Consumo insuficiente de la población: Desnutrición crónica</li> <li>▪ Déficit de infraestructura social: Población escolar excedente con relación a las aulas en uso; población que excede la capacidad de atención de las postas; población sin servicio de agua; población sin servicio de desagüe</li> <li>▪ Límites al desarrollo económico: Inaccesibilidad vial; población sin electrificación</li> </ul>
<b>FUENTE DE INFORMACIÓN</b>	Censo IX de Población y IV de Vivienda del año 1993	Censo IX de Población y IV de Vivienda del año 1993 y el Censo de Talla del año 1993	Censo IX de Población y IV de Vivienda del año 1993 y el Censo de Talla del año 1993	Censo IX de Población y IV de Vivienda de 1993 y la Encuesta Nacional de Hogares del IV Trimestre de 1997	Censo IX de Población y IV de Vivienda del año 1993; Censo de Talla del año 1993; Estadísticas del Ministerio de Educación 1998 y 1999, Encuestas y Estadísticas del Ministerio de Salud, Estadísticas de Transportes y Comunicaciones del año 1996, información sobre población beneficiaria de FONCODES durante el período 1993-1999.

3. Mapa de la Estrategia Focalizada de Lucha contra la Pobreza Extrema del Ministerio de la Presidencia (PROFINES)

Este mapa se elabora en dos etapas<sup>13</sup>:

- En primer lugar, se efectúa un ordenamiento a nivel distrital, en base al indicador de FONCODES. La variante del mapa distrital de PROFINES respecto del de FONCODES consiste en que 207 de los 419 distritos priorizados son seleccionados utilizando un criterio “demográfico”, es decir, incorporando la variable *tamaño poblacional* del distrito. Los 237 distritos restantes son seleccionados tomando en consideración exclusivamente el nivel relativo del indicador de FONCODES.
- En segundo lugar, se efectúa un ordenamiento por zonas censales (para el caso de distritos urbanos grandes) y centros poblados (para el caso de distritos más pequeños y rurales). Este ordenamiento se efectúa en base a la metodología del mapa de Necesidades Básicas Insatisfechas del INEI.

Las fuentes de información utilizadas son el Censo IX de Población y IV de Vivienda del año 1993 y el Censo de Talla del año 1993.

***Mapas recientemente propuestos, aun no operativos***

A pesar de no ser operativos en la actualidad, consideramos pertinente mencionar también dos mapas que se planea sean utilizados en el futuro inmediato por el sector público. (Ver Tabla 1)

1. Mapa del Ministerio de Economía y Finanzas

Este mapa presenta un ranking distrital elaborado en función de varios indicadores:

- El déficit de consumo de la población pobre
- El nivel de desigualdad en la distribución del gasto
- La severidad de la pobreza
- Los montos de inversión en cada distrito durante el período 1993-1999 en programas de lucha contra la pobreza.

Las fuentes de información en las que se basa son el Censo IX de Población y IV de Vivienda de 1993 y la Encuesta Nacional de Hogares del IV Trimestre de 1997.

2. Mapa del FONCODES 2000

Este mapa construye un índice sintético de pobreza y clasifica los distritos en 5 estratos (pobre extremo, muy pobre, pobre, regular y aceptable).

Los indicadores utilizados son:

- Indicadores que reflejan el consumo insuficiente de la población:
  - Tasa de desnutrición crónica
- Indicadores que reflejan déficit de infraestructura social:
  - Porcentaje de población escolar excedente en relación a las aulas en uso
  - Porcentaje de población que excede la capacidad de atención de las postas
  - Porcentaje de población sin servicio de agua
  - Porcentaje de población sin servicio de desagüe
- Indicadores que reflejan límites al desarrollo económico

---

13 En base a Bernedo (1999: 24).

- Inaccesibilidad vial
- Porcentaje de población sin electrificación

Las fuentes de información en las que se basa son el Censo IX de Población y IV de Vivienda del año 1993; Censo de Talla del año 1993; Estadísticas del Ministerio de Educación 1998 y 1999, Encuestas y Estadísticas del Ministerio de Salud, Estadísticas de Transportes y Comunicaciones del año 1996, información sobre población beneficiaria de FONCODES durante el período 1993-1999.

## 2.2.2 Avances y Limitaciones de los mapas en el Perú

Como fue señalado anteriormente, los mapas de pobreza presentan diversas limitaciones a nivel conceptual y operativo que deben ser tomadas en cuenta al momento de seleccionar y aplicar una metodología particular.

Al respecto, las metodologías de elaboración de mapas de pobreza utilizadas en el Perú difieren en diversos aspectos metodológicos como los métodos de identificación de la población en situación de pobreza, los métodos de agregación de los indicadores utilizados en la priorización de unidades geográficas, las fuentes de información utilizadas y la manera de integrarlas, entre otros. A continuación, se presenta una revisión crítica de los aspectos metodológicos que se consideran centrales para efectos del presente documento.

1. **Niveles de agregación geográfica.** Como se indicó en la sección anterior, diversos autores han señalado que la ganancia derivada de la utilización de instrumentos de focalización geográfica (en particular de mapas) es significativa únicamente para niveles de agregación “pequeños”. En el caso peruano, todos los mapas que se usan actualmente proporcionan criterios de priorización de gasto a nivel distrital<sup>14</sup>. Cabe notar que por debajo de este nivel de agregación solo se encuentran los centros poblados, los cuales, como se mostrará en la sección posterior, no permiten niveles de precisión aceptables para estimaciones de incidencia, brecha o severidad de pobreza salvo en aquellos centros poblados de tamaño grande.
2. **Diferencias en los métodos de identificación de hogares en situación de pobreza.** Al respecto, cabe recordar que existen fundamentalmente dos propuestas metodológicas en la formulación de instrumentos de focalización geográfica: el método de necesidades básicas insatisfechas (que enfoca su atención en variables observables que se refieren a aspectos estructurales de la pobreza) y el método de línea de pobreza (que constituye la base para la formulación de indicadores de incidencia, brecha y severidad de la pobreza, comúnmente utilizados). Mientras que la primera propuesta metodológica no requiere, en principio, ningún sistema de ponderación pues sólo identifica carencias (más aún basta una carencia para ser considerado pobre), en el segundo método las ponderaciones utilizadas en la construcción del indicador tienen como correlato teórico la función de utilidad indirecta con amplio sustento en la teoría económica (ver punto 3).

14 Vale la pena enfatizar que las metodologías de elaboración de mapas de pobreza han experimentado mejoras en precisión significativas a lo largo del tiempo. Shack (1996) señala, en relación a los mapas elaborados por FONCODES, que “..., en 1991 y 1992 se asignaron los recursos a nivel regional, conscientes de la necesidad de definir una unidad geográfica menor que la región, (en el año 1993) se determinó una asignación referencial de recursos a nivel provincial.”. Los criterios de asignación se basaron en indicadores provinciales hasta el año 1995.

En el caso peruano, la mayoría de mapas de pobreza han sido elaborados haciendo uso de indicadores construidos en base a estas dos propuestas metodológicas. Por un lado, el mapa de pobreza de NBI del INEI (1994) es un ejemplo de la primera opción metodológica; mientras que aquél elaborado por Shack y Díaz (2000) se basa, en principio, en la segunda opción<sup>15</sup>.

Complementariamente a los dos métodos mencionados, en los últimos años se ha propuesto la elaboración de indicadores compuestos, que pretenden integrar ambas metodologías con el fin de diferenciar la profundidad y persistencia de la pobreza. Un ejemplo de la opción "híbrida", particularmente importante en el Perú, lo constituye el indicador integrado utilizado por el INEI que pretende diferenciar a la población en situación de pobreza crónica, inercial y transitoria. Sin embargo, este indicador no es utilizado actualmente en la elaboración de mapas de pobreza.

Adicionalmente a estas metodologías de identificación, FONCODES ha elaborado mapas de pobreza (el vigente fue elaborado el año 1996 – el elaborado en el 2000 se encuentra en revisión –) en base a una metodología de identificación alternativa. Se seleccionan indicadores de manera *ad hoc* (inaccesibilidad vial, tasa de inasistencia escolar, tasa de desnutrición, entre otros) que son agregados haciendo uso de ponderaciones seleccionadas arbitrariamente. En base a esta agregación se obtiene indicadores que son estandarizados de manera que permitan elaborar un ordenamiento de pobreza de los agregados geográficos determinados (departamentos, provincias, distritos). Si bien con este sistema de ponderación se pretende establecer cierta diferenciación entre deficiencias registradas al interior y entre agregados geográficos, es necesario advertir que la arbitrariedad con la que se seleccionan estos ponderadores no permite saber qué se está midiendo realmente.

- 3. Diferencias en los métodos de agregación de los indicadores de pobreza.** Este es probablemente el tema más complicado en relación con la elaboración de mapas de pobreza, en tanto las técnicas de agregación incorporan complejidades específicas al método de identificación elegido; en particular, en el caso del método de línea de pobreza, el criterio de evaluación de la desigualdad al interior del agregado geográfico constituye un aspecto de especial relevancia (Sen,1997).

En el caso del mapa de pobreza del MEF (2000) que es elaborado en base al método de identificación de línea de pobreza, la técnica de agregación utilizada incorpora indicadores de severidad de la pobreza y desigualdad en la distribución de la misma al interior del agregado geográfico. Al respecto cabe analizar separadamente dos aspectos:

- La técnica de agregación utilizada se resume en la siguiente fórmula:

---

<sup>15</sup> “En principio” porque, como se mencionará más adelante, el indicador utilizado como base para elaborar los coeficientes de asignación presupuestal de este mapa, incorpora indicadores elaborados en función a la estimación de nivel de gasto (brecha, severidad y desigualdad).

$$A_i = \text{Déficit de Consumo Distrital}_i * \text{ITS distrital}_i * \text{Severidad Distrital}_i$$

$$\text{Indice de Pobreza} = \frac{A_i}{\sum_{i=1}^n A_i}$$

Donde ITS es el índice de Theil estandarizado y la severidad distrital se mide utilizando el indicador FGT (2). Si bien este indicador de pobreza tiene la cualidad de incorporar un elemento de medición de la desigualdad (elemento central en el proceso de agregación geográfica), tiene a su vez la limitación de no ser válido para todos los agregados geográficos. De acuerdo a los criterios de asignación que se derivan de esta metodología, una localidad con niveles de desigualdad nulos y tasas de pobreza cercanas al 100% no recibiría presupuesto destinado al combate de la pobreza, al ubicarse en el grupo de las localidades con menor índice de “pobreza” del país.

Al respecto, cabe notar que, entre las técnicas de agregación de indicadores de incidencia y desigualdad en un indicador de pobreza, la propuesta más reconocida es la de Sen (1997) que busca independizar los efectos de severidad y desigualdad:

$$\text{Indicador de Sen} = HI + H(1+I)Gp$$

Donde H es la tasa de incidencia (porcentaje de pobres en relación al total de habitantes de la localidad), I es el indicador de la brecha existente entre el nivel de ingresos y la línea de pobreza, y Gp es el coeficiente de Gini. Cabe señalar que este indicador puede ser ajustado, con algunas restricciones, para incorporar indicadores de desigualdad distintos al coeficiente de Gini.

- Adicionalmente, es necesario enfatizar que, incluso en caso de efectuar ajustes a la técnica de agregación utilizada, de manera que sea válida para agregados geográficos de diversas características, la metodología utilizada por el MEF conlleva, como es natural en este tipo de metodología, errores de estimación que afectan la precisión de los indicadores,. Estos errores se incrementan conforme se reduce el nivel de agregación, por lo que es necesario establecer los niveles para los cuales estos estimados son válidos<sup>16</sup>.

**4. Utilización de fuentes de información complementarias.** Uno de los retos más importantes que se enfrenta al momento de instrumentalizar la focalización geográfica consiste en desarrollar metodologías que permitan identificar áreas de dimensión reducida donde la concentración de pobres sea mayor (departamentos, provincias, distritos y de ser posible centros poblados), poniendo a su vez, a disposición de los usuarios, mecanismos de actualización permanente. Para ello, es necesario idear técnicas de integración de información no censal a la censal<sup>17</sup>.

16 Al respecto, cabe mencionar que los errores de filtración y subcobertura, medidos como los errores en la predicción al interior de la muestra de la encuesta de hogares, ha sido ampliamente explorado en la literatura especializada. El cálculo de los errores de estimación de tasas de pobreza a nivel de agregados geográficos serán discutidos en la siguiente sección.

17 Ver Bigman, Dercon, Guillaume y Lambotte (2000); Fofack (2000).

Al respecto, cabe notar que el mapa de FONCODES 1996 integró la información del Censo de Talla y del Censo de Población y Vivienda, y el mapa de pobreza de FONCODES 2000 estaría integrando, adicionalmente, información de fuentes tan diversas como estadísticas del Ministerio de Educación 1998 y 1999, encuestas y estadísticas del Ministerio de Salud, estadísticas de transportes y comunicaciones del año 1996, información sobre población beneficiaria de FONCODES durante el período 1993-1999.

El mapa de pobreza del MEF, por su lado, ha desarrollado una metodología que utiliza información de la Encuesta Nacional de Hogares de 1997 IV Trimestre además de la información censal de 1993.

Es importante notar que aunque es central incorporar la nueva información disponible para mejorar los mapas como instrumentos de focalización, la manera cómo esta es agregada es también importante. En el caso de FONCODES el esquema arbitrario de ponderación, impide tener una idea clara de cuál es el concepto de pobreza que subyace a la construcción del mapa.

### **3 MARCO METODOLÓGICO PARA LA ELABORACIÓN DE MAPAS DE POBREZA MEDIANTE LA UTILIZACIÓN COMPLEMENTARIA DEL CENSO DE POBLACIÓN Y VIVIENDA Y UNA ENCUESTA DE HOGARES SOBRE CONDICIONES DE VIDA Y POBREZA**

Como se indicó anteriormente, el objetivo de este documento es presentar un conjunto de aportes metodológicos (la mayor parte de ellos ya contemplados en la literatura especializada<sup>18</sup>), que mejoren la precisión con la que se estiman los indicadores de pobreza en el Perú a niveles reducidos de agregación geográfica. En particular, los indicadores elaborados en base al nivel de consumo o gasto de los hogares, que como se indicó en la sección anterior, cuenta con un correlato teórico de amplia aceptación.

La construcción de estos indicadores de bienestar a niveles reducidos de agregación geográfica como el distrital o el de centro poblado, requiere información sobre el consumo de cada uno de los hogares a nivel nacional. Sin embargo, la información censal disponible en la actualidad no incorpora información sobre consumo de los hogares; y las Encuestas de Hogares sobre Niveles de Vida y Pobreza<sup>19</sup> que incorporan esta información, son representativas únicamente a niveles de agregación geográfica elevados (i.e. a nivel nacional, urbano – rural, dominios geográficos), a los cuales la utilidad de un mapa de pobreza sería nula. En este sentido, el marco metodológico presentado en esta sección se centra en la utilización complementaria de ambas fuentes de información, prestando especial atención a la mejora en la precisión de las estimaciones de pobreza y en el cálculo de la confiabilidad de las mismas a niveles diferenciados de agregación geográfica.

La estrategia metodológica global consiste en estimar una relación funcional entre el nivel de *gasto per cápita* de los hogares, reportado en las encuestas de hogares sobre niveles de vida y pobreza, y un conjunto de variables construidas a partir de los reportes incluidos tanto en la encuesta de hogares como en el censo de población y vivienda utilizados. Con base en los parámetros que se obtienen de esta estimación se procede a efectuar la interpolación de los niveles de gasto de los hogares censales, siendo posible de esta manera calcular distintos indicadores de pobreza como la incidencia, brecha y severidad de la pobreza a distintos niveles de agregación geográfica. En particular, la metodología presentada se centra en la estimación del indicador de incidencia o tasa de pobreza<sup>20</sup>.

Esta estrategia metodológica es propuesta por Hentschel et al. (1998, 2000) y complementada por Elbers et al.(2000). En la presente sección se desarrollan los principales aspectos conceptuales y metodológicos de esta propuesta, y se incorporan

---

18 La propuesta original sobre la que se basa el presente documento ha sido desarrollada por Hentschel et al. (1998, 2000) y complementada por Elbers et al. (2000). Sin embargo, aplicaciones prácticas de estrategias de utilización complementaria de censo y encuesta de hogares ha sido aplicada por Minot (1998); Alderman, et al. (2000); Robles (1997); Shack et al. (2000), entre otros.

19 En el Perú, se dispone de dos encuestas de hogares sobre niveles de vida y pobreza, la elaborada por el INEI en el cuarto trimestre de cada año (ENAH0 del IV Trimestre) y la elaborada por el Instituto Cuánto cada tres años (ENNIV).

20 Cabe indicar, sin embargo, que esta metodología puede ser adaptada para efectuar estimaciones más complejas como las de brecha y severidad de la pobreza.

recomendaciones adicionales, ideadas a partir de la exploración conceptual y aplicación práctica de dicha metodología para el caso peruano.

El marco metodológico se divide en dos etapas:

- a. La primera etapa consiste en la estimación de un indicador de pobreza a nivel de hogares censales.
- b. La segunda etapa consiste en la agregación de este indicador por hogar a niveles decrecientes de agregación geográfica.

A continuación se describen ambas etapas del marco metodológico. Cabe indicar que la siguiente sección del documento detalla los principales resultados de una aplicación práctica de tal metodología, hecha en base a la Encuesta Nacional de Hogares 1998 del IV Trimestre y el 5% del Censo IX de Población y IV de Vivienda de 1993.

### **3.1 Estimación del Nivel de Gasto per cápita o de la Probabilidad de ser pobre del Hogar Censal**

La primera etapa tiene como objetivo obtener una estimación del nivel de bienestar de cada uno de los hogares de la base censal. Esta estimación, obtenida a nivel de hogar, servirá de base para los cálculos del indicador de pobreza que se utilizará para la elaboración del mapa en la segunda etapa. Al respecto, cabe resaltar que la unidad de análisis en esta primera etapa es el hogar, en tanto la segunda etapa pretende ofrecer estimaciones a nivel de individuos. El esquema seguido en esta primera etapa puede observarse en la Figura 1.

Esta primera etapa se resume en tres sub-etapas:

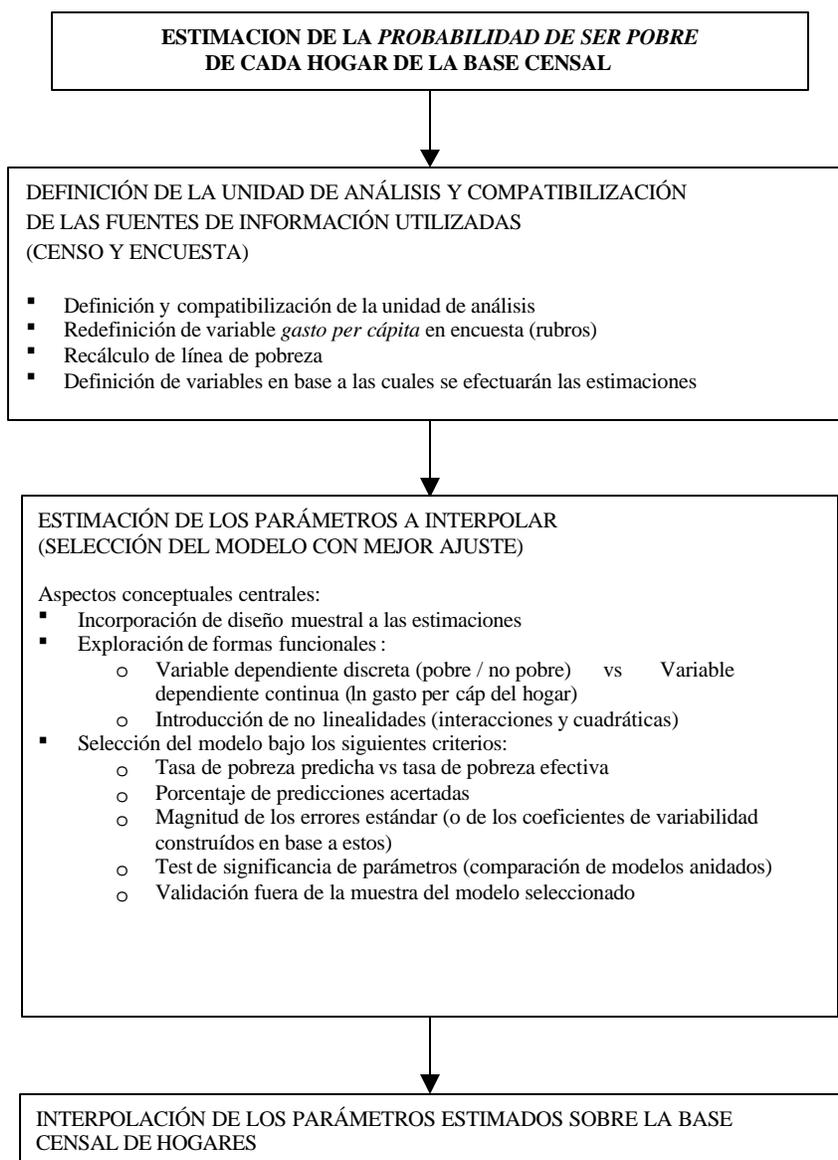
1. Definición de la unidad de análisis y compatibilización de las fuentes de información utilizadas (el censo y la encuesta).
2. Estimación de los parámetros que resumen la relación funcional entre las variables socio económicas y demográficas y el nivel de *gasto per cápita* o la *probabilidad de ser pobre* de los hogares muestreados en la encuesta.
3. Interpolación de los parámetros estimados en el paso anterior sobre la base de datos de los hogares censales.

A continuación, se detallan los aspectos centrales asociados a cada una de las sub-etapas mencionadas.

#### **3.1.1 Definición de la unidad de análisis y compatibilización de las fuentes de información utilizadas (censo y encuesta)**

Debido a que la metodología desarrollada busca integrar la información proporcionada por dos fuentes complementarias (encuesta de hogares y censo de población y vivienda), es necesario asegurar la compatibilidad de las mismas. En particular, la unidad de análisis constituye el elemento central con base en el cual se efectuará la compatibilización de la información incorporada en ambas bases de datos.

**Figura 1**  
**Esquema del Marco Metodológico– Etapa I**



### **Definición y compatibilización de la unidad de análisis**

El hogar constituye la unidad de análisis por excelencia en las mediciones de pobreza. En esta línea, se parte del reconocimiento del hogar como unidad de análisis para efectuar las estimaciones de pobreza para la construcción del mapa.

La definición de hogar alude a la definición de la composición del mismo, es decir, los criterios con base en los cuales se establece que un individuo es miembro o no del hogar en el que reside. Los reportes de las bases de datos utilizadas pueden ser agrupados a partir de diversas definiciones de hogar. En particular, la base censal disponible en la actualidad (censo 1993) no permite distinguir entre residentes habituales del hogar y otros individuos que se encuentran *de paso* en el hogar en el momento del Censo. En el

caso de la encuesta de hogares, por el contrario, es posible agrupar los reportes bajo definiciones alternativas de hogar, en tanto se incluyen reportes más detallados sobre el tipo de relación de parentesco y el tiempo de residencia en el hogar. En este contexto, el primer paso en la compatibilización de las bases de datos consiste en definir y compatibilizar la unidad de análisis más adecuada para efectuar las estimaciones del nivel de bienestar de los hogares y de diversos agregados geográficos.

La selección de la definición de hogar en base a la cual se efectuará la compatibilización de la información de las fuentes y las estimaciones respectivas debe considerar dos criterios (ver Anexo 1 para mayores detalles).

- a. La definición de hogar debe ser conceptualmente compatible con el diseño muestral de la encuesta, en particular, con la utilización de pesos muestrales. Los pesos muestrales reflejan la probabilidad de ser encuestado de un hogar y, en consecuencia, la probabilidad de ser encuestado de cada uno de sus miembros. Al respecto, la caracterización de los miembros del hogar debe garantizar que cada uno de los individuos de la población pertenece a un solo hogar, y evitar por tanto la duplicación de pesos muestrales que sería conceptualmente inconsistente, y empíricamente contradictoria con la expansión poblacional de los reportes de la encuesta<sup>21</sup>.

Por ejemplo, una encuesta de hogares recoge el reporte de tres clases de individuos en una vivienda, un “miembro presente” de la familia (por ejemplo, un hijo de 10 años), un “miembro ausente” de la familia (por ejemplo, un hijo de 13 años que se encuentra estudiando en otra ciudad bajo la tutela de un amigo de los padres), y un “residente habitual” de la vivienda (por ejemplo, un individuo que no pertenece a la familia, pero vive en la vivienda y comparte los alimentos con los demás “miembros presentes” de la misma). Al respecto, se podría optar por alguna de las siguientes definiciones de hogar, (i) todos los individuos que pertenecen a la familia, estén o no presentes, se consideran miembros del hogar; alternativamente, (ii) todos los individuos que residen habitualmente en el hogar, sean o no miembros de la familia, se consideran miembros del mismo.

En el primer caso, se excluye al “residente habitual” que no es miembro de la familia, pero es incluido (conceptualmente) en otro hogar, del cual es parte como “miembro ausente”. En el segundo caso, se excluye al “miembro ausente” del hogar, y éste pasa a ser miembro del hogar en el que reside (en el caso del niño de 13 años, pasaría a ser miembro del hogar del amigo de sus padres, con el que vive); y el “residente habitual” no miembro de la familia, forma parte de este hogar, y ya no del hogar al que pertenecen los demás “miembros presentes” de su familia. En caso contrario, es decir, en caso de definir como miembros de un hogar a las tres clases de individuos, se incurriría en la duplicación del “miembro ausente” de la familia y del “residente habitual” que no es miembro de la familia. A cada uno de ellos les correspondería dos probabilidades de ser encuestados, la que corresponde a los miembros de su familia y la que corresponde a los miembros de la familia con la que comparten vivienda y alimento.

---

21 Esta contradicción aparecería en la medida que las expansiones efectuadas a partir de los pesos muestrales deben permitir hacer estimaciones relativas a la población, y en caso de existir duplicaciones, tales expansiones excederían las proyecciones poblacionales oficiales.

- b. Por otro lado, la definición de la unidad de análisis debe permitir establecer con la mayor precisión posible el indicador de bienestar que será atribuido a los individuos que componen el hogar. Este aspecto es particularmente importante en el marco de esta metodología en tanto el indicador de bienestar utilizado se basa en el *gasto per cápita* del hogar, el cual es medido como la porción del gasto total del hogar que le corresponde a cada uno de los miembros que lo integran, parte que se asume equitativamente distribuída al interior del hogar.

Por ejemplo, es posible que la definición original de hogar incluya a un “residente habitual” que no es miembro de la familia y al cual se le provee exclusivamente de techo y alimento. Es muy probable que este residente tenga un nivel de vida distinto al de los demás miembros de la familia, por lo que sería impreciso atribuirle el nivel de vida del hogar con el que comparte únicamente techo y alimento.

### **Compatibilización de la información incorporada en las fuentes utilizadas**

Las variables que serán utilizadas en las estimaciones del nivel de bienestar de los hogares son construídas a nivel de hogar<sup>22</sup> en base a dos tipos de reporte:

- reportes por individuo, siempre que éste sea considerado un miembro del hogar (por ejemplo, el nivel educativo, la edad, la ocupación, entre otros); y
- reportes por hogar (como características de la vivienda en la que reside el hogar, los rubros de gasto en alimentos, electricidad, entre otros).

En general, estos reportes originales requerirán ser compatibilizados en ambas fuentes de datos, en tanto servirán de base en la construcción de las variables a ser utilizadas en las estimaciones<sup>23</sup>. En particular, es necesario compatibilizar:

- las categorías en las que se desagregan las variables originales, por ejemplo, el censo puede recoger reportes sobre la disponibilidad de alumbrado eléctrico en el hogar, y la encuesta puede desagregar esta variable en diversos tipos de alumbrado (eléctrico, kerosene, petróleo, vela, generador u otro); y
- la población sobre la cual se recogen los reportes de cada una de ellas. Por ejemplo, en el caso de reportes individuales, es posible que el censo presente reportes de nivel educativo de todos los individuos con 3 años o más, en tanto la encuesta restrinja sus reportes a los individuos de 6 años o más.

Adicionalmente a estos ajustes, en especial en caso de modificar la unidad de análisis original de las fuentes de información, es indispensable redefinir los rubros de gasto del hogar y recalcular el gasto que corresponde a los miembros del mismo, lo que implicará la redefinición de la línea de pobreza a utilizar en las estimaciones de los indicadores de bienestar<sup>24</sup>.

22 Ejemplos de estas variables se pueden encontrar en el Anexo 2, como es el caso de tamaño del hogar, porcentaje de miembros de 14 años y más que tienen nivel educativo superior en el hogar, gasto per cápita del hogar (variable que se pretende estimar sobre la base censal de hogares).

23 Al respecto, en el Anexo 2 se presenta la comparación de las variables originalmente reportadas en el censo y en la encuesta de hogares, utilizadas en el ejercicio realizado.

24 En el Anexo 1 se presenta en detalle los ajustes efectuados a ambas variables en el marco del ejercicio realizado.

### **3.1.2 Estimación de los parámetros que resumen la relación funcional entre las variables socio económicas y demográficas y el nivel de *gasto per cápita* o la *probabilidad de ser pobre* de los hogares muestreados en la encuesta**

Una vez compatibilizadas las unidades de análisis y las variables a ser utilizadas en la elaboración del mapa, se procede a la estimación de los parámetros que expresan la relación funcional entre las características de los hogares y los niveles de consumo que estos reportan en la encuesta de hogares. La estimación de estos parámetros requiere, como se detallará a continuación, de la exploración de formas funcionales alternativas apropiadas para la estimación de indicadores de bienestar -considerando el diseño muestral utilizado en la preparación y aplicación de la encuesta-, así como de la definición de criterios de selección, y de la validación del modelo con mejor ajuste que será utilizado en la interpolación.

A continuación se desarrollan algunas de las ideas principales que sustentan la utilización de las especificaciones metodológicas desarrolladas en el presente documento para la estimación de los parámetros y que permitirán la estimación de los niveles de bienestar de los hogares de la base censal.

#### **Incorporación del diseño muestral**

La incorporación del diseño muestral atraviesa el proceso de estimación de parámetros y selección del modelo con mejor ajuste. La importancia de incorporar especificaciones del diseño muestral de la encuesta de hogares utilizada se delinea a continuación.

La muestra de las encuestas de hogares sobre niveles de vida y pobreza realizadas en el Perú es, típicamente, probabilística, estratificada, multietápica y de áreas<sup>25</sup>. Normalmente, las estimaciones efectuadas en base a encuestas con diseños muestrales de este tipo son ajustadas únicamente tomando en cuenta lo que se denomina pesos muestrales o factores de expansión dejando de lado otros elementos como la especificación de las unidades de muestreo o la estratificación utilizada. Estos pesos indican la probabilidad de ser encuestado del hogar particular al que se hace referencia (más específicamente, de la vivienda en la cual éste reside, esto debido a que el marco muestral es realizado en base a parámetros de ubicación geográfica).

Si bien el ajuste derivado de la incorporación de los pesos muestrales o factores de expansión, permite la obtención de coeficientes insesgados, no permite la aproximación correcta a los errores estándar e intervalos de confianza de tales estimadores, para lo que es necesario incorporar los demás elementos que conforman el diseño muestral<sup>26</sup>. Comúnmente, la aplicación exclusiva de los factores de expansión, es decir, la exclusión de otros elementos del diseño muestral, conduce a una subestimación importante de los errores estándar de los estimadores. Ello, a su vez, ocasiona que variables que no

25 Respecto del diseño de la ENAHO 1998 del IV Trimestre (utilizada en el ejemplo desarrollado en la siguiente subsección) ver Medina (1999) - Manual de Procedimientos para el Diseño y Selección de la Muestra de la ENAHO 98 (Documentos Metodológicos, Programa para el Mejoramiento de las Encuestas y la Medición de las Condiciones de Vida en América Latina y el Caribe (MECOVI)) -.

26 El error estándar de un estimador es una medida de las desviaciones que obtendríamos si efectuáramos consecutivas estimaciones sobre muestras diferentes elegidas tomando el mismo diseño muestral que el aplicado a la muestra sobre la que efectuamos la estimación. Stata, [U] 26 -236-

contribuyen de manera importante a explicar la varianza total del *gasto per cápita* sean erróneamente incluidas en la estimación del modelo.

Es importante reconocer que la pertinencia de la incorporación del diseño muestral en el análisis de regresión está sujeta a controversia tanto a nivel teórico como práctico. La discusión básicamente se centra en dos aspectos: i) incluir o no el diseño muestral en la estimación de los coeficientes. ii) corregir o no los errores estándar asociados a dichos coeficientes. (Deaton,1997; Pfeffermann,1993). A pesar de la discusión teórica existente, debemos mencionar que la inclusión o no de los factores de ponderación y demás elementos del diseño muestral depende de que esperamos de una regresión. Si las regresiones son básicamente descriptivas deberíamos incorporar el diseño muestral. Básicamente en estudios como el presente donde lo que se busca es mejorar el ajuste para determinado agregado poblacional (por ejemplo un dominio geográfico) más que estimar relaciones causales a nivel de la unidad de análisis (por ejemplo un hogar), hay consenso en torno a la necesidad de corregir los errores estándar del modelo estimado (Deaton,1997).

Los elementos del diseño muestral que se toman en cuenta para obtener estimadores insesgados y conocer sus errores estándar, son los siguientes:

- 1. Las unidades de muestreo**, que son conglomerados o aglomeraciones de población, cuyo tamaño es definido en la etapa de elaboración del marco muestral de la encuesta. Estas aglomeraciones son definidas a niveles decrecientes de agregación geográfica, por ejemplo, en el caso de la encuesta de hogares utilizada en el presente documento se observan unidades de muestreo a tres niveles: las unidades primarias (UPM), secundarias (USM) y terciarias (UTM) de muestreo.

Siguiendo con el ejemplo (ENAH 1998 del IV Trimestre), la selección de estas unidades de muestreo sigue varias etapas. En una primera etapa, se efectúa una selección sistemática, proporcional al tamaño, de las UPM; y en una segunda etapa, se efectúa una selección, también sistemática y proporcional al tamaño, de las USM que se ubican al interior de las UPM seleccionadas. Al interior de estas unidades (USM), se selecciona aleatoriamente un número determinado de viviendas (UTM) a ser encuestadas (se encuesta a la totalidad de hogares que se encuentran en el interior de las viviendas muestreadas). Este proceso de muestreo multietápico permite, debido a la mayor homogeneidad al interior de un conglomerado que entre conglomerados, mejorar la representatividad de las observaciones, mediante la selección de un mayor número de conglomerados y un reducido número de hogares por conglomerado.

- 2. La estratificación**, que refiere a la clasificación de UPM en los denominados *estratos*, con el fin de reducir la heterogeneidad de características socio económicas entre UPM.

En el caso de la encuesta de hogares utilizada, se definen cinco estratos, tres urbanos y dos rurales, sobre los cuales se efectúa la selección de UPM referida anteriormente. Con ello se pretende asegurar la representatividad de cada estrato en la muestra encuestada.

**3. Los pesos muestrales o factores de expansión.** Si tenemos un proceso de muestreo con reemplazo que involucra dos etapas podemos definir  $p_c$  como la probabilidad de selección del conglomerado  $c$  y  $p_{ic}$  como la probabilidad de seleccionar el hogar  $i$ , dentro del conglomerado  $c$ , en la segunda etapa. Finalmente podemos escribir el peso de la observación  $i$  en la muestra como:

$$w_{ic} = \frac{1}{p_c p_{ic}}$$

Cabe mencionar que en el caso de encuestas de hogares  $p_c$  incorpora la probabilidad de que tanto la UPM como USM donde se encuentra el hogar sean seleccionadas.

### Exploración de formas funcionales

La forma funcional que se utiliza para modelar la relación entre el nivel de bienestar (medido por el logaritmo del *gasto per cápita* de los hogares o por la condición de pobre o no pobre de los hogares) y las características de los hogares puede afectar las estimaciones de tasa de pobreza de manera importante. En este sentido, la exploración de formas funcionales debe tomar en consideración la capacidad de las mismas de producir estimadores insesgados del indicador de incidencia de la pobreza.

La forma funcional usualmente utilizada en la elaboración de mapas de pobreza que aplican metodologías similares a la desarrollada en este documento es la ecuación lineal de gasto, que estima el logaritmo de *gasto per cápita* de los hogares. En base a este estimador, es posible efectuar cálculos de indicadores de pobreza como el de incidencia a partir del conteo de hogares predichos como pobres (aquellos cuya predicción de gasto no supera la *línea de pobreza*). Sin embargo, como señala Hentschel et al. (1998, 2000), el conteo de pobres genera un estimador sesgado de la tasa de pobreza, resultando más apropiado efectuar las estimaciones de tasa de pobreza sobre estimaciones de la probabilidad de un hogar de ser pobre<sup>27</sup>. Esta probabilidad viene dada por la función de distribución acumulada normal estándar, evaluada en la diferencia entre la *línea de pobreza* y el *gasto per cápita* estimado del hogar (estandarizada por el estimador de la desviación estándar del error del modelo).

$$\hat{P}_i = f\left(\frac{\ln z - X_i' \hat{\mathbf{b}}}{\hat{\mathbf{s}}}\right)$$

Donde  $f$  representa una distribución acumulada normal estándar,  $X_i$  representa el vector de características reportadas por el hogar  $i$ ,  $\hat{\mathbf{b}}$  representa el vector de parámetros estimados en el modelo lineal (logarítmico),  $\hat{\mathbf{s}}$  es el estimador de la desviación estándar del error del modelo lineal (logarítmico), y  $\ln z$  es el logaritmo natural de la *línea de pobreza* del agregado geográfico al que pertenece el hogar  $i$ .

---

<sup>27</sup> Como indican Hentschel et al. (2000: 155), es recomendable estimar la probabilidad de ser pobre de cada hogar, tomando en cuenta que el modelo no explica con absoluta precisión el nivel de consumo del hogar (el R2 nunca es igual a 1), y que el consumo predicho está basado en datos muestrales.

Alternativamente al modelo lineal (logarítmico), es posible efectuar estimaciones a partir de un modelo probit, el cual se orienta a la estimación de la probabilidad de un hogar de ser pobre, en base a la cual es posible obtener estimadores insesgados de la tasa de pobreza.

Además de comparar los resultados entre un modelo lineal (logarítmico) de gasto y un modelo probit, es importante notar las potenciales ganancias derivadas de incorporar no-linealidades en el gasto y en la probabilidad de que un hogar sea pobre. Por ejemplo, la tasa de pobreza puede no crecer a una tasa constante al incrementarse el número de miembros de un hogar, o disminuir proporcionalmente, con los años de educación. En este sentido, es útil explorar modelos alternativos que incorporen estas relaciones no lineales entre las variables consideradas originalmente y el nivel de bienestar de los hogares.

En base a lo anteriormente expuesto, es posible afirmar que se dispone de, por lo menos, cuatro modelos alternativos que permitirían efectuar estimaciones de la probabilidad de un hogar de ser pobre sobre la base censal.

### **Criterios de selección del modelo con mejor ajuste**

Una vez definidos los modelos pertinentes para la estimación de la *probabilidad de ser pobre* de un hogar particular, es necesario proceder a una comparación sistemática de las ventajas y desventajas de cada uno, con el fin de seleccionar el modelo con mejor ajuste que será utilizado en la interpolación. Al respecto, algunos criterios considerados centrales en este proceso de selección son:

1. Evaluar un indicador que refleje la distancia entre la tasa de pobreza estimada en base a los datos de la encuesta (“tasa de pobreza efectiva”) y la tasa de pobreza predicha en base a cada uno de los modelos estimados.

Al respecto, es necesario tomar en cuenta el error estándar de la “tasa de pobreza efectiva” (calculada en base a los reportes de la encuesta), que permite conocer el intervalo de confianza de la misma. En caso de obtener diferencias que no excedan el intervalo de confianza referido, este criterio no debe ser tomado en cuenta en la selección del modelo, en tanto las diferencias observadas no serían estadísticamente significativas.

2. Evaluar un indicador de acierto de la predicción de la situación de pobreza de los hogares. Este indicador se evalúa a nivel de cada uno de los hogares particulares de la encuesta. Es decir, se efectúa una predicción de la situación de pobreza de un hogar y se le compara con la situación real (reportada por el hogar en la encuesta).

Este indicador puede ser complementado con un análisis de la tasa de filtración y subcobertura. En particular, puede considerarse preferible minimizar alguna de estas tasas; por ejemplo, la tasa de subcobertura. En este caso, el modelo con menor tasa de subcobertura podría ser superior a otro con una mayor tasa de acierto, cuyos desaciertos se encuentren concentrados en hogares pobres predichos por el modelo como no pobres.

3. Evaluar la magnitud del error estándar de las estimaciones de tasa de pobreza de cada uno de los modelos. Es posible, en su defecto, efectuar comparaciones del coeficiente de variabilidad de cada una de las estimaciones.

El error estándar de las estimaciones de pobreza efectuadas en base a cada uno de los modelos utilizando la base de datos de la encuesta (estimaciones intra muestra), es la raíz cuadrada de la varianza del estimador, la cual se calcula de la siguiente manera<sup>28</sup>:

$$Var(\bar{P}) = \sum_i^N \frac{\bar{P}_i(1-\bar{P}_i)}{N^2} + \left( \frac{\partial \bar{P}}{\partial \mathbf{b}} \right)' Var(\mathbf{b}) \left( \frac{\partial \bar{P}}{\partial \mathbf{b}} \right) + \left( \frac{\partial \bar{P}}{\partial \mathbf{s}^2} \right)^2 \frac{2\mathbf{s}^4}{n-k-1}$$

Donde:

*En el modelo lineal (logarítmico):*

$$\frac{\partial \bar{P}}{\partial \mathbf{b}_j} = \sum_i^N \frac{1}{N} \left( \frac{-x_{ij}}{\mathbf{s}} \right) f \left( \frac{\ln z - X'_i \mathbf{b}}{\mathbf{s}} \right)$$

$$\frac{\partial \bar{P}}{\partial \mathbf{s}^2} = -\frac{1}{2} \sum_i^N \frac{1}{N} \left( \frac{\ln z - X'_i \mathbf{b}}{\mathbf{s}^3} \right) f \left( \frac{\ln z - X'_i \mathbf{b}}{\mathbf{s}} \right)$$

*En el modelo probit (asumiendo que la varianza es asintóticamente igual a 1):*

$$\frac{\partial \bar{P}}{\partial \mathbf{b}_j} = \sum_i^N \frac{1}{N} (x_{ij}) \hat{P}_i$$

$$\frac{\partial \bar{P}}{\partial \mathbf{s}^2} = 0$$

Siendo:

$\hat{P}_i$  la *Probabilidad de ser pobre* del hogar i, predicha por el modelo sobre la base de la encuesta

$\bar{P}$  la *Tasa de Pobreza* de un agregado geográfico particular (estimada a partir de las predicciones por hogar)

28 Las fórmulas presentadas a continuación constituyen una variante de las propuestas por Hentschel et al. (1998; 2000), y que serán utilizadas para la estimación del error estándar de las estimaciones de tasa de pobreza poblacional efectuadas sobre la base censal. Cabe resaltar que la utilización de esta variante se debe a que las fórmulas originales refieren a una unidad de análisis poblacional (individuo), pertinente en la segunda etapa de este marco metodológico; requiriéndose en esta primera etapa una adecuación que refiera a la unidad de análisis pertinente: el hogar.

- $s$  el estimador de la desviación estándar del error del modelo
- $b$  el vector de  $k$  parámetros estimado en el modelo
- $k$  el número de variables independientes del modelo (variables explicativas del modelo)
- $x_{ij}$  la característica  $j$  reportada por el hogar  $i$  de la encuesta
- $N$  el número total de hogares del agregado geográfico determinado
- $n$  el número de reportes por hogar con base en los cuales se estima el modelo

Más adelante se desarrolla en detalle el significado de cada uno de los componentes del error de predicción. Cabe adelantar, sin embargo, que el primer sumando representa el error idiosincrático y los dos últimos el error del modelo, derivado del diseño muestral y la forma funcional asumidas en la estimación del modelo en cuestión.

4. Test de significancia de parámetros. Este test permite efectuar comparaciones entre modelos anidados, sin embargo, no es útil para comparar formas funcionales distintas (probit respecto al lineal (logarítmico)) o modelos no anidados.

Cabe señalar, además, que en caso de incorporar el diseño muestral de la encuesta en la estimación de los modelos, el test de significancia apropiado es el Test de Wald ajustado por el diseño muestral<sup>29</sup>.

5. Validación de la predicción del modelo con mejor ajuste fuera de la muestra, es decir, evaluar el ajuste del modelo sobre los reportes de otra encuesta de hogares. En caso de no ser validado, sería necesario evaluar el ajuste del segundo mejor fuera de la muestra.

Asimismo, en caso que los cuatro criterios anteriores no hayan permitido establecer la superioridad clara de un modelo respecto de los demás, es recomendable efectuar este ejercicio de validación para los que hayan mostrado ser igualmente buenos, y seleccionar aquél con mejor ajuste fuera de la muestra.

### 3.1.3 Interpolación de los parámetros estimados en 3.1.2 sobre la base censal de hogares

La interpolación de los parámetros estimados en 3.1.2 sobre la base censal de hogares constituye el último paso de esta primera etapa y permite obtener la estimación del nivel de bienestar de los hogares de la base censal (en el caso de haber seleccionado un modelo lineal (logarítmico) o un modelo probit, permite obtener un estimador de la probabilidad del hogar de ser pobre).

Al respecto, es necesario tomar en consideración dos tipos de ajuste que permiten evitar la pérdida innecesaria de observaciones sobre las cuales se efectúe la interpolación:

---

<sup>29</sup> Al respecto, es útil revisar las opciones que el programa econométrico utilizado en la estimación de los modelos ofrece. En particular, STATA ofrece la opción del Test de Wald ajustado por el diseño muestral, particularmente útil en el caso de diseños muestrales con menos de 100 UPM muestreadas. Mayor detalle puede encontrarse en <http://www.stata.com/support/faqs/stat/lrtest.html>.

1. Un potencial problema al utilizar la base censal de hogares es el de la ausencia de reportes de determinadas variables que forman parte del modelo estimado en la subetapa anterior. Al respecto, se considera recomendable efectuar imputaciones siempre que estas no generen sesgos sistemáticos.

La media y la mediana son comúnmente utilizadas para efectuar imputaciones. Al respecto, se recomienda la utilización de la mediana del centro poblado al que pertenece el hogar que requiere la imputación, con el fin de evitar que ésta sea afectada por valores extremos.

Adicionalmente, se recomienda evaluar que las imputaciones efectuadas no generen sesgos sistemáticos. En caso de verificar que las imputaciones generan sesgos sistemáticos, es necesario identificar las observaciones (hogares) que, habiendo sido imputadas, son responsables de tales sesgos y eliminarlas de la base sobre la cual se pretende efectuar la interpolación.

2. Es indispensable homogeneizar los rangos de variación de las variables sobre las cuales se efectúa la interpolación. Es decir, es necesario eliminar las observaciones (hogares) de la base censal que presenten reportes que exceden el rango de variación de sus equivalentes en la encuesta. Esto permite garantizar la validez de la interpolación, debido a que no es posible inferir el nivel de bienestar de hogares que presentan reportes que escapan el rango observado en la encuesta y que sirvió para modelar la relación entre las características de los hogares y su nivel de bienestar (relación reflejada por los parámetros interpolados).

En caso de no ser posible efectuar la interpolación de algunas observaciones por cualquiera de estas dos razones (i.e. no fue posible efectuar imputaciones, o los rangos de variación de las variables pertinentes excedieron los observados en la encuesta), es indispensable tomar nota del tamaño de los hogares perdidos con el fin de generar un factor de ajuste a nivel de centros poblados que permita recuperar el tamaño del hogar perdido. De esta manera, se garantiza la estimación de la tasa de pobreza sobre el tamaño poblacional efectivo del agregado geográfico de interés (centro poblado, distrito, provincia o departamento).

### **3.2 Agregación de las Estimaciones por Hogar: Tasa de Pobreza y Error de Predicción a Niveles Diferenciados de Agregación Geográfica**

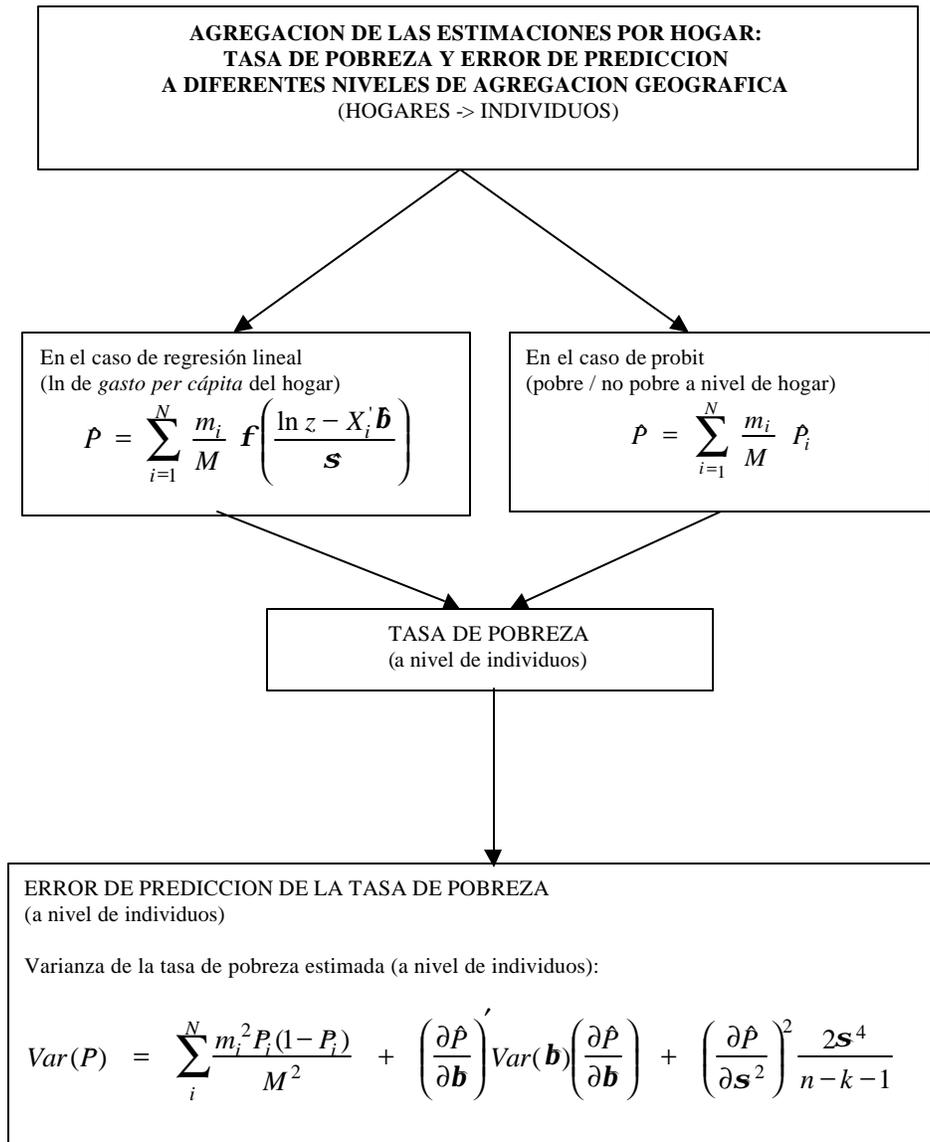
La segunda etapa de la metodología propuesta consiste en obtener estimaciones de la tasa de pobreza a niveles de agregación geográfica diferentes, en particular, a nivel de departamento, provincia, distrito y centro poblado. Para ello, se requiere agregar las estimaciones obtenidas en la primera etapa. El esquema seguido en esta etapa puede observarse en la Figura 2.

Antes de presentar las fórmulas con base en las cuales se efectuará esta agregación, cabe resaltar que los cálculos efectuados en la etapa anterior se hicieron exclusivamente a nivel de hogar. Como es obvio, los indicadores de tasa de pobreza que interesan en la elaboración del mapa son indicadores a nivel de individuos, es decir, indicadores que tomen en consideración las diferencias en el tamaño de los hogares pobres y no pobres

al interior de los agregados geográficos. Por ello, se incorporará como ponderador el tamaño del hogar en las estimaciones de tasa de pobreza y error de predicción.

**Figura 2**

**Esquema del Marco Metodológico - Etapa II**



### 3.2.1 Estimación de la Tasa de Pobreza a niveles geográficos determinados

Como se mencionó anteriormente, las posibilidades de cálculo de tasas de pobreza varían en función del tipo de variable predicha en la etapa de interpolación.

A partir del modelo lineal (logarítmico) es posible efectuar agregaciones derivadas de la variable *gasto per cápita del hogar*, en tanto ésta es la variable predicha sobre la base censal. Como se indicó anteriormente, estas agregaciones pueden efectuarse de dos maneras:

- Calculando la tasa de pobreza por medio de un conteo de pobres. Es decir, se calcula el número de individuos cuyo *gasto per cápita* no cubre el equivalente monetario de la canasta básica de consumo. Como se indicó anteriormente, esta opción no es la óptima, en tanto produce estimadores sesgados.
- Efectuando estimaciones de la probabilidad de que un hogar sea pobre, por medio de la función de distribución acumulada normal estándar, evaluada en la diferencia entre la *línea de pobreza* y el *gasto per cápita* estimado del hogar (estandarizada por el estimador de la desviación estándar del error). Así, la *probabilidad de ser pobre* de la localidad en cuestión (incidencia de pobreza) puede ser calculada como la probabilidad promedio de los hogares que residen al interior, ponderando cada uno de los mismos por el número de miembros.

En el caso del modelo probit, la variable predicha es la *probabilidad de que un hogar sea pobre*, por lo que el cálculo de la incidencia de la pobreza en una localidad particular se reduce al cálculo de la probabilidad promedio de ser pobre del conjunto de hogares que residen en dicha localidad.

De esta manera, la tasa de pobreza estimada sobre la base censal a un nivel de agregación geográfica determinado ( $\hat{p}$ ) estaría dada por:

En el modelo lineal (logarítmico):

$$\hat{p} = \sum_{i=1}^N \frac{m_i}{M} f\left(\frac{\ln z - X_i' \hat{\mathbf{b}}}{\mathbf{s}}\right)$$

En el modelo probit:

$$\hat{p} = \sum_{i=1}^N \frac{m_i}{M} \hat{p}_i$$

Donde:

$f$  representa una distribución acumulada normal estándar,  $X_i$  representa el vector de características reportadas por el hogar  $i$  en la base censal,  $\hat{\mathbf{b}}$  representa el vector de parámetros estimados en el modelo lineal (logarítmico),  $\mathbf{s}$  es el estimador de la desviación estándar del error del modelo lineal (logarítmico), y  $\ln z$  es el logaritmo natural de la *línea de pobreza* del agregado geográfico al que pertenece el hogar  $i$ .

$\hat{p}_i$  representa la probabilidad del hogar  $i$  de ser pobre estimada en el modelo probit.  
 $M$  es el tamaño del agregado geográfico particular (población total) y  $m_i$  representa el número de miembros del hogar  $i$ .  
 $N$  es el número de hogares de la base censal.

### **3.2.2 Cálculo del error estándar de la tasa de pobreza estimada a niveles geográficos determinados**

Finalmente, se calcula el error estándar de las estimaciones de tasa de pobreza al nivel de agregación geográfica particular.

Como se señaló anteriormente, las estimaciones de pobreza derivadas de metodologías que involucren estimaciones e imputaciones intermedias, requieren ser complementadas con el cálculo de los intervalos de confianza de las mismas. Esto se debe a que el proceso de estimación del indicador involucra fuentes de error diversas que pueden limitar la confiabilidad del mismo y con ello su utilidad práctica.

De acuerdo a Elbers et al. (2000) la precisión con la que se estima los indicadores de pobreza dependen de tres tipos de errores:

- Un error idiosincrático, asociado a variables no contempladas en el modelo estimado que generan desviaciones entre el gasto efectivo (no observado en la base censal) y el valor esperado de acuerdo al modelo. El componente idiosincrático de la varianza crece cuando uno focaliza su atención en poblaciones más pequeñas (que se alejan del patrón medio establecido a niveles mayores de agregación), y por tanto es el factor que limita la capacidad de desagregación. Cabe señalar que la rapidez con que se produzca el incremento del error conforme se reduce el tamaño poblacional de la localidad en cuestión dependerá del poder explicativo de las variables exógenas y en consecuencia de la importancia del componente idiosincrático del modelo.
- Un error asociado al modelo que incorpora a su vez los errores derivados del marco muestral utilizado en la encuesta de hogares y de la forma funcional asumida. Para una localidad determinada, el error del modelo dependerá también de la distancia entre los valores que toman las variables explicativas de los hogares censales que residan en ella y los que tomen los hogares muestreados en la encuesta. Al respecto, la metodología presentada en la sub-sección anterior pretende minimizar este error a través de la incorporación del marco muestral y la mejora en la especificación del modelo para minimizar los errores de filtración y subcobertura. Es importante anotar que la magnitud de este error dependerá de cuán bueno sea el ajuste de la ecuación a ser interpolada, y no del nivel de desagregación geográfica al que se pretenda efectuar la interpolación (y con ella, la predicción de la tasa de pobreza).
- Un error de cómputo, derivado de la necesidad de obtener, en algunos casos, un cálculo aproximado y no exacto del indicador de pobreza. Esta tercera fuente de error aparece cuando no es posible derivar analíticamente los dos primeros componentes, siendo necesario efectuar aproximaciones numéricas para obtener una medida de los mismos. Este es el caso de indicadores como el de severidad de la

pobreza (FGT2) <sup>30</sup>. En este sentido, la técnica de cómputo que permite aproximar la varianza del estimador de pobreza conlleva un error que se denomina error de cómputo.

Cabe indicar que el indicador de incidencia o tasa de pobreza no requiere de aproximaciones numéricas, por lo que la varianza se resume en los dos componentes anteriormente expuestos.

El error de predicción es la raíz cuadrada de la varianza del estimador de tasa de pobreza, la cual viene dada por (Hentschel et al. 2000: 157):

$$Var(\hat{P}) = \sum_i^N \frac{m_i^2 \hat{P}_i (1 - \hat{P}_i)}{M^2} + \left( \frac{\partial \hat{P}}{\partial \mathbf{b}} \right)' Var(\mathbf{b}) \left( \frac{\partial \hat{P}}{\partial \mathbf{b}} \right) + \left( \frac{\partial \hat{P}}{\partial \mathbf{s}^2} \right)^2 \frac{2\mathbf{s}^4}{n - k - 1}$$

Donde:

*En el modelo lineal (logarítmico):*

$$\frac{\partial \hat{P}}{\partial \mathbf{b}_j} = \sum_i^N \frac{m_i}{M} \left( \frac{-x_{ij}}{\mathbf{s}} \right) f \left( \frac{\ln z - X_i' \mathbf{b}}{\mathbf{s}} \right)$$

$$\frac{\partial \hat{P}}{\partial \mathbf{s}^2} = -\frac{1}{2} \sum_i^N \frac{m_i}{M} \left( \frac{\ln z - X_i' \mathbf{b}}{\mathbf{s}^3} \right) f \left( \frac{\ln z - X_i' \mathbf{b}}{\mathbf{s}} \right)$$

*En el modelo probit (asumiendo que la varianza es asintóticamente igual a 1):*

$$\frac{\partial \hat{P}}{\partial \mathbf{b}_j} = \sum_i^N \frac{m_i}{M} (x_{ij}) \hat{P}_i$$

$$\frac{\partial \hat{P}}{\partial \mathbf{s}^2} = 0$$

Siendo:

$\hat{P}_i$  la Probabilidad de ser pobre del hogar  $i$ , estimada en la interpolación sobre la base censal de hogares.

$\hat{P}$  la Tasa de Pobreza de un agregado geográfico particular (estimada a partir de la interpolación censal)

---

30 Al respecto, ver Elbers et al. (2000: 6).

- $N$  el número total de hogares del agregado geográfico determinado
- $x_{ij}$  la característica  $j$  reportada por el hogar  $i$
- $s$  el estimador de la desviación estándar del error del modelo
- $b$  el vector de parámetros estimados en el modelo
- $n$  el número de reportes por hogar de la encuesta
- $k$  el número de variables independientes del modelo (variables socio económicas y demográficas que permiten estimar el nivel de bienestar de los hogares de la encuesta)

El primer sumando representa el error idiosincrático, mientras que los dos últimos componentes representan el error del modelo, tanto aquél que afecta la precisión con la que se han estimado los parámetros del modelo usado en la interpolación, como la precisión con la que se ha estimado la varianza de los errores en esa primera etapa.

#### **4 APLICACIÓN PRÁCTICA DE LA METODOLOGÍA DE ESTIMACIÓN DE INDICADORES DE BIENESTAR A NIVELES REDUCIDOS DE AGREGACIÓN GEOGRÁFICA: PRINCIPALES RESULTADOS**

En la presente sección se presenta los principales resultados de una aplicación práctica de la metodología desarrollada en la sección anterior. Cabe resaltar que los resultados que serán presentados no son generalizables a otros ejercicios prácticos. Es decir, la selección de la unidad de análisis o del modelo a utilizar, y obviamente las estimaciones de tasa de pobreza o error de predicción, se supedita a las bases de datos particulares utilizadas en este ejercicio.

El ejercicio práctico que se resume a continuación se desarrolló en base a la utilización de dos fuentes de información:

- La Encuesta Nacional de Hogares (ENAH) 1998 del IV Trimestre, la cual presenta 6,952 reportes por hogar, representativos a nivel de dominio (Costa Urbana, Costa Rural, Sierra Urbana, Sierra Rural, Selva Urbana, Selva Rural, Lima Metropolitana)<sup>31</sup>.
- Una muestra sistemática (seleccionada por departamento) del 5% del Censo de IX Población y IV de Vivienda realizado en 1993, la cual presenta 238,248 reportes a nivel de hogar (aproximadamente 1.1 millón de reportes individuales a nivel nacional).

La presentación de los resultados sigue en gran medida el esquema utilizado en la presentación del marco metodológico. Cabe notar que, por consideraciones de claridad expositiva, algunos de los resultados no son presentados en detalle en esta sección, sin embargo, pueden encontrarse en los Anexos 1, 2 y 3.

##### **4.1 Etapa I: Estimación del Nivel de Gasto per cápita o de la Probabilidad de ser pobre del Hogar Censal**

Siguiendo la metodología detallada en la sección anterior, esta primera etapa fue orientada a la estimación del nivel de bienestar de los hogares de la base censal. Debido a que se exploraron dos modelos lineales *logarítmicos* y dos modelos *probit*, el indicador de bienestar obtenido en la interpolación fue el estimado de la *probabilidad de ser pobre* de los hogares censales. Cabe recordar que ambos tipos de modelo proporcionan un estimador de la *probabilidad de ser pobre*; el lineal *logarítmico* a partir de la distribución acumulada normal estándar evaluada en la diferencia entre la tasa de pobreza y la predicción de gasto (estandarizada por la varianza estimada del error), y el *probit* a partir de la predicción directa de la *probabilidad de ser pobre* del hogar.

---

31 Cabe notar que esta metodología es aplicable a cualquier otra encuesta sobre condiciones de vida y pobreza, sin embargo, para efectos de este documento se optó por la ENAH 1998 del IV trimestre en consideración a su amplia cobertura en términos comparativos a la ENNIV 1997. En la primera se contaba con 6,952 hogares, en tanto en la segunda se disponía de 3,843 hogares.

A continuación se presentan los resultados obtenidos en el proceso de compatibilización de la unidad de análisis, la selección del modelo con mejor ajuste y la interpolación censal.

#### **4.1.1 Definición de la unidad de análisis y compatibilización de las fuentes de información utilizadas (censo y encuesta)**

Los reportes del censo y de la encuesta son tomados a tres niveles de agregación: vivienda, hogar e individuo. Al respecto, existe un problema de compatibilidad entre las definiciones de hogar censal y las que aparecen en los documentos de la encuesta de hogares. En tanto la unidad de análisis para la elaboración del mapa de pobreza es el hogar, el primer paso consiste en compatibilizar las definiciones de hogar que serán utilizadas para la generación de variables a incluir en las regresiones y posteriormente en la interpolación.

La base censal no permite redefinir la composición de los hogares, por lo que las posibilidades de redefinición de la composición de los mismos fue posible únicamente a nivel de la base de datos de la encuesta ENAHO. Se reformularon las definiciones de hogar presentadas en los documentos de la encuesta de manera que no dejaran de cumplir con las condiciones expuestas en la sección anterior.

Estas dos definiciones (ambas conceptualmente válidas) son:

- Hogar Familiar (definición original de la encuesta). Constituido por el conjunto de personas, sean o no parientes, que ocupan en su totalidad o en parte una vivienda, comparten las comidas principales y atienden en común otras necesidades vitales. Se incluye a personas a quienes el jefe de hogar considera que son miembros del hogar por razones de afecto (ahijados, compadres, padrinos, etc.).
- Hogar ENAHO (definición reformulada). Constituido por los residentes habituales del hogar, es decir, (i) miembros del hogar familiar que residen habitualmente en él (se excluye a los que se encuentran ausentes 30 días o más), y (ii) individuos que no son miembros del hogar familiar pero residen habitualmente en él (están presentes hace 30 días o más).

Al respecto, cabe indicar que se decidió trabajar con la definición de hogar familiar, debido a que refleja de manera más precisa el *gasto per cápita* que estaría beneficiando efectivamente a los miembros del hogar. En el Anexo 1 se puede observar con mayor detalle el análisis efectuado para seleccionar la definición de hogar más adecuada para procesar los reportes de la encuesta.

Es importante hacer notar que este cambio de definición condujo a modificaciones en el nivel de *gasto per cápita* de los hogares de la encuesta y en las líneas de pobreza utilizadas en la construcción de los indicadores de pobreza. En particular, el cambio en las tasas de pobreza estimadas a partir de los reportes de la encuesta (respecto de las calculadas por el INEI), tomando en consideración tales modificaciones, se debió fundamentalmente al ajuste efectuado sobre las líneas. Cabe notar, sin embargo que este ajuste en la línea se debió más a los otros cambios realizados (requerimientos

calóricos mínimos y coeficiente de Engel utilizado) que al cambio en la definición de gasto. A continuación se presentan las diferencias entre las tasas de pobreza estimadas tomando en consideración los ajustes referidos y las tasas de pobreza oficiales estimadas por el INEI<sup>32</sup>.

**Cuadro 1**

Dominios	Tasa de Pobreza Extrema		Tasa de Pobreza (Extr y No Extr)	
	Ajustada	INEI	Ajustada	INEI
Costa Urbana	6.5%	5.1%	42.2%	27.9%
Costa Rural	22.0%	19.9%	67.7%	43.1%
Sierra Urbana	9.9%	8.6%	37.3%	32.8%
Sierra Rural	45.7%	42.9%	80.2%	61.9%
Selva Urbana	8.7%	6.7%	44.2%	33.5%
Selva Rural	28.7%	26.2%	79.2%	53.2%
Lima Metropolitana	2.6%	2.0%	24.6%	21.2%
Nacional	17.3%	15.6%	49.4%	37.3%

Fuente: elaboración propia en base a la ENAHO 98-4

Adicionalmente, se efectuaron ajustes en la definición de las variables originalmente incluídas en ambas bases de datos con el fin de compatibilizarlas (en términos de categorías y población sobre la cual se recogieron los reportes en cada una). Asimismo, se definió un conjunto de variables, construidas en base a las primeras, que fueron utilizadas en las estimaciones de los parámetros a interpolar. Al respecto, el Anexo 2 presenta el detalle correspondiente.

#### **4.1.2 Estimación de los parámetros que resumen la relación funcional entre las variables socio económicas y demográficas y el nivel de *gasto per cápita* o la *probabilidad de ser pobre* de los hogares muestreados en la encuesta**

##### **Modelación de las relaciones funcionales entre nivel de bienestar y características socio económicas y demográficas de los hogares**

Se definió un conjunto de 68 variables con base en las cuales se modelaron las relaciones funcionales a estimar para predecir el nivel de bienestar de los hogares. Estas 68 variables (que se denominarán *variables lineales*) fueron construidas a nivel de hogar, en base a las variables simultáneamente reportadas en la encuesta y el censo<sup>33</sup>. En el Cuadro 2 se presenta el listado de estas 68 variables, las cuales reflejan diversas características socio económicas y demográficas de los hogares, algunas de ellas construidas a partir de la agregación de características individuales de los miembros del

32 Cabe indicar que estas tasas de pobreza fueron calculadas en base a la totalidad de reportes de la encuesta (6,952). Estas tasas de pobreza no coinciden exactamente con las presentadas en el análisis de las estimaciones y la interpolación censal, debido a que estas últimas se efectúan en base a un conjunto de observaciones más reducido (6,913) que excluye las observaciones extremas de la encuesta.

33 Como se indicó en 4.1.1, previamente se compatibilizó la definición y población sobre las cuales se recogieron tales reportes en la encuesta y el censo. Con ello se aseguró que las variables construidas a partir de los reportes originales sean compatibles entre ambas bases.

hogar (nivel educativo, edad, género, ocupación), y otras construidas a partir del reporte del hogar (características de la vivienda).

**CUADRO 2**

<b>VARIABLES CONSIDERADAS EN LA FORMULACIÓN DE LOS MODELOS</b>	
<u>Nombre</u>	<u>Definición de la variable</u>
miemhog	número de miembros del hogar
tothog01	número de hogares que residen en la vivienda
electric	el hogar tiene alumbrado eléctrico
habingr	el hogar utiliza algún espacio de la vivienda para actividad que genere ingresos
radio	el hogar tiene radio
tvblnegr	el hogar tiene tv blanco y negro
tvcolor	el hogar tiene tv a color
refrig	el hogar tiene refrigeradora / congeladora
maqcoser	el hogar tiene máquina de coser
eqsonido	el hogar tiene equipo de sonido
videogr	el hogar tiene video grabadora
lavadora	el hogar tiene lavadora
biciclet	el hogar tiene bicicleta
autocam	el hogar tiene auto o camioneta
triciclo	el hogar tiene triciclo
motocicl	el hogar tiene motocicleta
comput	el hogar tiene computadora
ntrabhog	número de trabajadores del hogar
npension	número de pensionistas del hogar
sexjef	sexo del jefe de hogar
pmiemuj	proporción de mujeres respecto del total de miembros del hogar, incluyendo al jefe de hogar
edadjef	edad del jefe de hogar
pmuj0a13	proporción de mujeres de 0 a 13 años respecto del total de miembros, incluyendo al jefe de hogar
phom0a13	proporción de hombres de 0 a 13 años respecto del total de miembros, incluyendo al jefe de hogar
pmuj65	proporción de mujeres mayores de 65 años respecto del total de miembros, incluye al jefe de hogar
phom65	proporción de hombres mayores de 65 años respecto del total de miembros, incluye al jefe de hogar
panalfab	proporción de analfabetos de 14 años o +, incluyendo al jefe de hogar
pasisten	proporción de miembros de 5 años o + que asisten actualmente a algún centro de enseñanza (respecto del total de miembros de 5 años y +), incluyendo al jefe de hogar
edusinj	jefe del hogar sin nivel educativo (sin nivel o con ed inicial)
edjpri nc	jefe del hogar con educación primaria incompleta
edjprcom	jefe del hogar con educación primaria completa
edjseinc	jefe del hogar con educación secundaria incompleta
edjsecom	jefe del hogar con educación secundaria completa
edjsnuin	jefe del hogar con educación superior no universitaria incompleta
edjsnuco	jefe del hogar con educación superior no universitaria completa
edjsuinc	jefe del hogar con educación superior universitaria incompleta
pmedusin	proporción de miembros del hogar de 5 años y + sin nivel educativo (sin nivel o con ed inicial), excluyendo al jefe de hogar, respecto del total de miembros del hogar
pmedupri	proporción de miembros del hogar de 5 años y + con ed primaria incompleta, excluyendo al jefe de hogar, respecto del total de miembros del hogar
pmedprco	proporción de miembros del hogar de 5 años y + con ed primaria completa, excluyendo al jefe

	de hogar, respecto del total de miembros del hogar"
pmedsein	proporción de miembros del hogar de 5 años y + con ed secundaria incompleta, excluyendo al jefe de hogar, respecto del total de miembros del hogar
pmedseco	proporción de miembros del hogar de 5 años y + con ed secundaria completa, excluyendo al jefe de hogar, respecto del total de miembros del hogar
pmesnuin	proporción de miembros del hogar de 5 años y + con ed superior no universitaria incompleta, excluyendo al jefe de hogar, respecto del total de miembros del hogar
pmesnuco	proporción de miembros del hogar de 5 años y + con ed superior no universitaria completa, excluyendo al jefe de hogar, respecto del total de miembros del hogar
pmesuinc	proporción de miembros del hogar de 5 años y + con ed superior universitaria incompleta, excluyendo al jefe de hogar, respecto del total de miembros del hogar
pmesucom	proporción de miembros del hogar de 5 años y + con ed superior universitaria completa, excluyendo al jefe de hogar, respecto del total de miembros del hogar
pmieocum	proporción de mujeres ocupadas de 14 años y + respecto del total de miembros del hogar, incluyendo al jefe de hogar
pcotal	proporción de miembros, incluyendo al jefe de hogar, cuya ocupación es la de ejecutivos, profesionales, técnicos, oficinistas, servicios, otros (FFAA y policiales)
pagrotot	proporción de miembros de 14 años y +, incluyendo al jefe de hogar, que trabajan en un negocio agrícola, pecuario, de silvicultura o caza, respecto del total de miembros ocupados del hogar
pcomtot	proporción de miembros de 14 años y +, incluyendo al jefe de hogar, que trabajan en un negocio comercial, respecto del total de miembros ocupados del hogar
vvalquil	la vivienda es alquilada
vvpropia	la vivienda es propia - totalmente pagada o comprándola a plazos
numhab	número de habitaciones de la vivienda, son contar con baño, cocina, pasadizos
prladril	las paredes de la vivienda son de ladrillo o bloque de cemento
prpiedra	las paredes de la vivienda son de piedra o sillar con cal o cemento
pradobe	las paredes de la vivienda son de adobe o tapia
prquinch	las paredes de la vivienda son de quincha
prbarro	las paredes de la vivienda son de piedra con barro
prmadera	las paredes de la vivienda son de madera
prestera	las paredes de la vivienda son de estera
parquet	el material predominante en los pisos es parquet o madera pulida
lamina	el material predominante en los pisos es lámina asfáltica, vinílico o similares
losetas	el material predominante en los pisos es loseta, terrazos o similares
madera	el material predominante en los pisos es madera
cemento	el material predominante en los pisos es cemento
tierra	el material predominante en los pisos es tierra
aguaenvv	el abastecimiento de agua procede de red pública dentro de la vivienda
higenvv	el servicio higiénico está conectado a red pública dentro de la vivienda
telef	el hogar tiene teléfono

Adicionalmente, se generaron variables que permitieran expresar las interacciones entre pares de *variables lineales*, con el fin de introducir no linealidades en la modelación del *gasto per cápita* o de la *probabilidad de ser pobre* (el Anexo 3 detalla las variables individuales y las interacciones utilizadas en los modelos para cada uno de los dominios de la ENAHO).

Con base en este conjunto de variables se exploraron las siguientes formas funcionales<sup>34</sup>:

- Dos modelos lineales logarítmicos con el fin de modelar la relación funcional lineal entre un grupo de características del hogar y el logaritmo de *gasto per cápita* asociado al mismo (uno de los modelos incluye solo *variables lineales*, mientras que el otro incorpora, adicionalmente, la interacción entre las mismas).

34 Estas estimaciones, así como la construcción de los indicadores utilizados en el proceso de selección del modelo con mejor ajuste, son efectuadas sobre la base de datos sin incluir observaciones extremas.

- Dos modelos probit que permiten modelar la relación funcional entre un grupo de características del hogar y la *probabilidad de ser pobre* del mismo (uno de los modelos incluye solo *variables lineales*, mientras que el otro incorpora, adicionalmente, la interacción entre las mismas).

A continuación se muestra una breve descripción de las características de los modelos estimados.

**Cuadro No. 3**

Modelo	Observaciones	Número de Variables		Test de Wald ajustado (por diseño muestral)*	Goodness-of-fit**
		Lineales***	Interacciones		
<b>Lineal Sin Interacciones</b>					
Costa Urbana	1157	35	-	F(33, 131) = 193.19	0.564
Costa Rural	632	22	-	F(20, 270) = 97.13	0.577
Sierra Urbana	1328	32	-	F(31, 133) = 186.65	0.664
Sierra Rural	1609	33	-	F(30, 260) = 58.47	0.540
Selva Urbana	588	30	-	F(28, 136) = 1367.4	0.642
Selva Rural	571	22	-	F(19, 271) = 31.56	0.470
Lima Metropolitana	790	32	-	F(29, 580) = 106.62	0.741
<b>Lineal Con Interacciones</b>					
Costa Urbana	1157	35	63	F(55, 109) = 1985.88	0.669
Costa Rural	632	22	31	F(46, 244) = 25220.42	0.646
Sierra Urbana	1328	32	94	F(70, 94) = 567.14	0.797
Sierra Rural	1609	33	66	F(94, 196) = 43237.24	0.612
Selva Urbana	588	30	53	F(35, 129) = 800.57	0.754
Selva Rural	571	22	21	F(33, 257) = 467.26	0.516
Lima Metropolitana	790	32	53	F(79, 530) = 275.1	0.825
<b>Probit Sin Interacciones</b>					
Costa Urbana	1157	21	-	F(20, 144) = 21.44	0.789
Costa Rural	655	20	-	F(19, 271) = 12.29	0.787
Sierra Urbana	1328	22	-	F(21, 143) = 24.97	0.850
Sierra Rural	1609	23	-	F(22, 268) = 13.55	0.841
Selva Urbana	588	21	-	F(20, 144) = 54.12	0.841
Selva Rural	579	23	-	F(22, 268) = 9.96	0.825
Lima Metropolitana	729	16	-	F(15, 602) = 12.47	0.888
<b>Probit Con Interacciones</b>					
Costa Urbana	1157	21	22	F(42, 122) = 37.91	0.825
Costa Rural	655	20	9	F(28, 262) = 11.35	0.823
Sierra Urbana	1328	22	13	F(34, 130) = 15.13	0.866
Sierra Rural	1609	23	15	F(34, 130) = 15.13	0.850
Selva Urbana	588	21	9	F(29, 135) = 44.7	0.846
Selva Rural	579	23	7	F(29, 261) = 10.7	0.835
Lima Metropolitana	729	16	3	F(18, 599) = 12.25	0.888

\* Todos los F son significativos al 1%. La incorporación de diseño muestral restringe grados de libertad.

\*\* En el caso de modelos lineales se presenta el  $R^2$  como medida de ajuste. En el caso de modelos probit se presenta la tasa de aciertos (debido a que, al incorporar el diseño muestral los estimadores obtenidos no son estrictamente de máxima verosimilitud, por lo que no es adecuado utilizar el estimador de Mc Fadden como pseudo- $R^2$ ).

\*\*\* La constante estimada en el modelo se ha contabilizado como una variable lineal más.

Como se aprecia en el Cuadro 2, estas relaciones funcionales fueron modeladas separadamente para cada uno de los dominios geográficos de la encuesta. Cada una de estas estimaciones incorporó el diseño muestral de la encuesta. En particular, para el caso de los dominios Costa Urbana, Costa Rural, Sierra Urbana, Sierra Rural, Selva Urbana y Selva Rural, se incorporaron los pesos muestrales correspondientes a cada hogar y la especificación de las unidades primarias de muestreo y el estrato en los que se ubica la vivienda en la que reside cada hogar. Para el caso del dominio de Lima Metropolitana, debido a que constituye en sí mismo una unidad primaria de muestreo, se optó por utilizar, además de los pesos muestrales y los estratos, la especificación de las unidades secundarias de muestreo<sup>35</sup>.

### **Proceso de selección del modelo con mejor ajuste**

Una vez efectuadas las estimaciones, se procedió a efectuar la comparación de los cuatro modelos bajo diferentes criterios para seleccionar el modelo de mejor ajuste. Al respecto, cabe resaltar que, siguiendo la metodología indicada en la sección anterior, los indicadores con base en los cuales se efectuó la comparación se construyeron tomando en consideración el peso muestral o factor de expansión por hogar. Asimismo, las predicciones por hogar utilizadas en cada uno de los criterios se refieren a la probabilidad de un hogar de ser pobre.

Como se indicó en la sección anterior, en el caso de los modelos probit, la predicción del modelo consiste en la probabilidad de un hogar de ser pobre. Sin embargo, para el caso de los modelos lineales, es necesario efectuar la estimación de esta probabilidad a partir de la distribución acumulada normal estándar evaluada en la diferencia entre el logaritmo de la línea de pobreza del dominio al cual pertenece el hogar y la predicción del logaritmo de *gasto per cápita* del mismo (diferencia estandarizada por el estimador de la varianza del error del modelo). De esta manera, se garantiza que las estimaciones de tasa de pobreza, efectuadas en el marco de cada uno de los modelos, sean insesgadas.

A continuación se detalla el proceso de selección del modelo con mejor ajuste, en función de cada uno de los criterios propuestos en la Parte 3 del documento.

### **Criterios Para Seleccionar el Modelo de Mejor Ajuste**

#### **1. Distancia entre la tasa de pobreza estimada a partir de las predicciones del modelo y la tasa de pobreza estimada a partir de los reportes de la encuesta (denominada “tasa de pobreza efectiva”).**

El Cuadro 4 muestra las tasas de pobreza estimadas a partir de las predicciones de cada uno de los modelos, así como la tasa de pobreza estimada a partir de los reportes efectivos de los hogares muestreados. Las cifras en *negritas* indican el modelo que ofrece la estimación de la tasa de pobreza, por dominio, más cercana a la tasa de pobreza efectiva. Asimismo, se presenta el error estándar de la estimación de tasa de pobreza a partir de los reportes de la encuesta (la denominada “tasa de pobreza efectiva”).

---

35 Respecto del factor de corrección por poblaciones finitas, no fue incluido en las estimaciones debido a la multi-estratificación de la encuesta.

Al respecto, cabe indicar que las diferencias observadas a nivel de dominio entre la tasa de pobreza estimada por cada uno de los modelos y la tasa de pobreza efectiva no es estadísticamente significativa, en ninguno de los casos, en tanto no exceden el error estándar de las estimaciones de la tasa de pobreza efectiva. Por ello, este criterio no será considerado en la comparación entre modelos.

**Cuadro 4**

Dominios	Estimación a partir de reportes de la encuesta		Estimaciones a partir de predicciones de los modelos			
	Tasa de Pobreza	Error Estándar	Tasa de Pobreza Lineal sin interacciones	Tasa de Pobreza Lineal con interacciones	Tasa de Pobreza Probit sin interacciones	Tasa de Pobreza Probit con interacciones
Costa Urbana	35.60%	2.23%	34.22%	34.98%	<b>35.08%</b>	35.04%
Costa Rural	59.25%	4.12%	<b>58.87%</b>	58.72%	59.75%	<b>59.63%</b>
Sierra Urbana	30.55%	3.46%	32.59%	32.10%	29.87%	<b>29.99%</b>
Sierra Rural	74.15%	2.03%	<b>74.20%</b>	74.03%	73.96%	73.92%
Selva Urbana	36.83%	3.52%	33.76%	34.90%	36.16%	<b>36.35%</b>
Selva Rural	71.85%	2.83%	69.52%	69.77%	71.78%	<b>71.79%</b>
Lima Metrop	18.91%	1.80%	20.41%	20.36%	<b>18.87%</b>	18.80%

## 2. Indicador de aciertos individuales en la predicción de la condición de pobreza del hogar (Tasa de Acierto).

La tasa de acierto de cada uno de los modelos fue establecida a partir de la comparación entre la predicción de la *probabilidad de ser pobre* de los modelos y el reporte de condición de pobreza que aparece en la encuesta. Al respecto, cabe indicar que se consideró hogares predichos como pobres a aquellos hogares cuya predicción de *probabilidad de ser pobre* igualaba o superaba el 50%. Así, el número de hogares predichos correctamente, según este criterio, fue expandido por el peso muestral correspondiente y prorrateado entre el número de hogares para los cuales se efectuó alguna predicción. De esta manera, se obtuvo el porcentaje de aciertos de cada uno de los modelos, mostrado en el Cuadro 5.

**Cuadro 5**

Dominios	Porcentaje de aciertos			
	Lineal sin interacciones	Lineal con interacciones	Probit sin interacciones	Probit con interacciones
Costa Urbana	77.81%	<b>83.07%</b>	78.88%	82.50%
Costa Rural	76.90%	79.43%	78.66%	<b>82.29%</b>
Sierra Urbana	82.20%	84.83%	85.03%	<b>86.58%</b>
Sierra Rural	83.83%	<b>85.91%</b>	84.06%	84.99%
Selva Urbana	82.45%	<b>86.05%</b>	84.05%	84.55%
Selva Rural	82.03%	<b>83.80%</b>	82.51%	83.50%
Lima Metrop.	86.80%	87.82%	<b>88.78%</b>	88.75%
Nacional	82.80%	85.38%	84.31%	85.62%

En consecuencia, el criterio de comparación de tasa de acierto muestra que los modelos que incorporan no linealidades (lineal con interacciones y probit con interacciones) son superiores en ajuste a los que no las incorporan. La única excepción a esta conclusión es la tasa de acierto del modelo probit sin interacciones para el caso de Lima Metropolitana. Al respecto, sin embargo, la diferencia con la tasa de acierto del probit con interacciones no es significativa lo que no invalida lo concluido en base a las tasas de acierto de los otros dominios.

### 3. Comparación de los errores estándar de las estimaciones de pobreza de los modelos.

Un criterio adicional que se utilizó para identificar los modelos de mejor ajuste fue estimar la magnitud de los errores estándar de las tasas de pobreza calculadas en base a las predicciones intramuestra de cada uno de los modelos. Como se observa en la Cuadro 6 los errores estándar de las estimaciones de tasa de pobreza de los modelos lineales son sistemáticamente menores que los correspondientes a las tasas estimadas en el marco de modelos probit.

**Cuadro 6**

Dominios	Error estándar de las estimaciones de tasa de pobreza efectuadas a partir de predicciones de los modelos			
	Lineal sin interacciones	Lineal con interacciones	Probit sin interacciones	Probit con interacciones
Costa Urbana	2.13%	<b>1.99%</b>	2.84%	3.53%
Costa Rural	<b>4.16%</b>	4.88%	6.25%	7.27%
Sierra Urbana	2.72%	<b>2.18%</b>	2.47%	2.52%
Sierra Rural	<b>3.45%</b>	3.94%	6.66%	6.86%
Selva Urbana	<b>2.44%</b>	2.61%	4.18%	4.17%
Selva Rural	<b>4.71%</b>	5.50%	8.31%	9.09%
Lima Metrop	1.76%	<b>1.61%</b>	2.14%	2.24%

Cabe mencionar que adicionalmente, se efectuaron los cálculos del coeficiente de variabilidad de las estimaciones de cada modelo, y se obtuvieron resultados similares, es decir, los coeficientes de variabilidad de los modelos lineales fueron sistemáticamente menores que los correspondientes a los modelos probit<sup>36</sup>.

### 4. Pruebas de significancia de parámetros.

Adicionalmente, y buscando comparar modelos con interacciones y modelos sin interacciones, se utilizó el Test de Wald ajustado por diseño muestral. De esta manera se evaluó la significancia de los parámetros adicionales incorporados en los modelos con interacciones.

<sup>36</sup> Es importante mencionar que el error estandar del modelo probit va a tender asintóticamente a 1, aunque en muestras finitas no será efectivamente 1 (ver Davidson y MaKinnon, pp.524).

Como se muestra en el Cuadro 7, la incorporación de interacciones, tanto en el modelo lineal (logarítmico) como en el modelo probit, mejora el ajuste del modelo.

**Cuadro 7**

Dominios	Test de Wald ajustado por diseño muestral	
	Lineal con interacciones	Probit con interacciones
Costa Urbana	F(56,108) = 1731.17 Prob > F = 0.0000	F(22,142) = 12.26 Prob > F = 0.0000
Costa Rural	F(31,259) = 155.41 Prob > F = 0.0000	F(9,281) = 7.28 Prob > F = 0.0000
Sierra Urbana	F(71,93) = 1152.40 Prob > F = 0.0000	F(13,151) = 6.65 Prob > F = 0.0000
Sierra Rural	F(66,224) = 29.71 Prob > F = 0.0000	F(15,275) = 7.34 Prob > F = 0.0000
Selva Urbana	F(36,128) = 490.05 Prob > F = 0.0000	F(9,155) = 1.81 Prob > F = 0.0703
Selva Rural	F(21,269) = 103.93 Prob > F = 0.0000	F(7,283) = 3.61 Prob > F = 0.0010
Lima Metropolitana	F(53,556) = 20.31 Prob > F = 0.0000	F(3,614) = 5.89 Prob > F = 0.0006

### Balance de la comparación

En el marco de este ejercicio, debido a las diferencias estadísticamente no significativas entre la tasa de pobreza efectiva y las tasas de pobreza estimadas a partir de las predicciones intra muestra de cada uno de los modelos, el primer criterio no es tomado en cuenta en la selección del modelo con mejor ajuste.

Sin embargo, los tres criterios adicionales permiten definir a uno de los modelos como el modelo con mejor ajuste. Por un lado, y para escoger entre el modelo con y sin interacciones se encuentra que el indicador de aciertos inclinaría el balance a favor de los modelos que incorporan interacciones entre variables. Consistentemente, y bajo el criterio de superioridad estadística de ajuste modelo del modelo con interacciones respecto del modelo sin interacciones, el Test de Wald ajustado muestra que tanto en el caso del probit como en el caso del modelo lineal logarítmico, las interacciones son significativas.

Por otro lado, habiendo seleccionado los modelos con interacciones y buscando escoger entre el modelo probit y el modelo lineal logarítmico, se encontró bajo el criterio de los errores estándar que los modelos lineales presentan menores errores estándar en referencia a la estimación de tasas de pobreza.

A partir de estos criterios, por tanto, se podría indicar como el modelo a elegir el modelo lineal logarítmico con interacciones<sup>37</sup>.

El siguiente paso consistiría en efectuar el ejercicio de validación de los modelos que, de acuerdo a los criterios anteriores, prueben tener el mejor ajuste. Esta validación podría hacerse estimando los modelos sobre los reportes de la encuesta de hogares de Cuánto o sobre otra encuesta de hogares del INEL. Para efectos del ejercicio, sin embargo, no se ha realizado el ejercicio de validación. Se asumió que el modelo lineal logarítmico con interacciones fue validado a partir de alguna otra encuesta y, por consiguiente, la interpolación censal se efectuó en base a los parámetros estimados a partir de dicho modelo.

#### **4.1.3 Interpolación de los parámetros estimados en 4.1.2 sobre la base censal de hogares**

En esta sección se desarrolla el último paso de la Etapa I explicada en la Figura 1. Como se indicó al inicio de la sección, la interpolación censal se efectuó sobre el 5% de los reportes del Censo de Población y Vivienda de 1993.

Al respecto, cabe indicar que fue necesario efectuar imputaciones de las medianas del centro poblado al que pertenecían los hogares que no presentaban los respectivos reportes. Adicionalmente, debido a que no se disponía de la base censal completa, se presentaron algunos problemas de ausencia de reporte a nivel de centros poblados, por lo que se optó por efectuar (solo para estos casos) imputaciones de medianas del distrito. Cabe indicar, sin embargo que la probabilidad de que este tipo de imputaciones (medianas distritales) sea necesaria cuando se trabaje sobre la base censal completa es mínima.

Adicionalmente, se homogeneizó el rango de variación de la base censal respecto del observado en la base de la encuesta, para el caso de las variables del modelo lineal logarítmico con interacciones.

Asimismo, se generó un factor de ajuste por la pérdida de observaciones derivada de los dos ajustes anteriores. En el Anexo 1 se presenta en detalle cada uno de los ajustes referidos.

#### **4.2 Etapa II: Agregación de las Estimaciones por Hogar: Tasa de Pobreza y Error de Predicción a Niveles Diferenciados de Agregación Geográfica**

Con base en las estimaciones del logaritmo de *gasto per cápita* de los hogares censales, obtenidas en base a la interpolación de los parámetros estimados en el modelo lineal logarítmico con interacciones, fue posible efectuar las estimaciones de la tasa de pobreza y el cálculo del nivel de precisión de las mismas a diferentes niveles de

---

<sup>37</sup> Sin embargo, es importante reiterar que lo relevante es conocer los criterios de evaluación ya que el modelo específico que se escoja al final siempre va a depender de la información con la que se esté trabajando.

agregación geográfica. Para ello se utilizó la metodología descrita en la sección anterior (ver Figura 2 y sección 3.2).

A continuación se presentan las tasas de pobreza estimadas en base a la interpolación censal, el error estándar de las mismas, y la tasa de pobreza obtenida a partir de los reportes de la encuesta, a nivel de dominios geográficos.

Cabe recordar que las estimaciones que se presentan a continuación refieren al tamaño poblacional, y no al número de hogares. Por ello, las estimaciones efectuadas en 4.1.2 de la tasa de pobreza efectiva (obtenida en base a los reportes de la encuesta de hogares utilizada y expandida a nivel de hogares) difieren ligeramente de las presentadas a continuación.

**Cuadro 8**

Dominios	Estimación a partir de reportes de la Encuesta	Estimación a partir de la Interpolación Censal	
	Tasa de Pobreza	Tasa de Pobreza	Error Estándar
Costa Urbana	42.52%	50.35%	3.43%
Costa Rural	68.05%	76.71%	7.74%
Sierra Urbana	37.38%	51.96%	4.16%
Sierra Rural	80.39%	86.72%	6.14%
Selva Urbana	44.41%	53.33%	4.72%
Selva Rural	79.60%	82.34%	8.21%
Lima Metropolitana	24.65%	37.69%	2.77%

### Descomposición de la varianza de las estimaciones de pobreza

Como se detalló en la sección anterior, existen dos fuentes de error en la estimación de la tasa de pobreza. Por un lado, el error idiosincrático, asociado a variables no contempladas en el modelo estimado y que genera desviaciones entre el gasto efectivo (no observado en la base censal) y el valor esperado de acuerdo al modelo. Por otro lado, el error del modelo, derivado del diseño muestral de la encuesta y de la forma funcional utilizada en la estimación de los parámetros interpolados.

A continuación se presenta la descomposición de la varianza de la tasa de pobreza estimada a partir de la interpolación censal efectuada. El primer factor refleja la participación del error idiosincrático en la varianza referida. Los dos factores restantes representan la participación del error del modelo en la varianza de la tasa de pobreza estimada (el segundo asociado a la varianza de los parámetros estimados y el tercero asociado al estimador de la varianza de los errores -  $s$ ).

Como se desprende del Cuadro 9, el componente más importante de la varianza es el error del modelo, lo que reforzaría la idea de que existe un espacio importante de mejora de la predicción de la tasa de pobreza (exploración de modelos con mejor ajuste) si es

que se mantiene este patrón a niveles de desagregación más detallados. En tal sentido sería útil evaluar qué mejoras podrían incluirse. Por ejemplo, sería importante explorar qué otras variables deberían incluirse en las encuestas de hogares que permitan mejorar la eficiencia de la predicción. Otro ejercicio interesante consistiría en evaluar si es que el error del modelo se reduce en caso de utilizar la encuesta de hogares elaborada por el Instituto Cuánto (ENNIV), en vez de utilizar la ENAHO.

**Cuadro 9**

Dominios	Descomposición de la varianza del estimador de Tasa de Pobreza de los dominios		
	Error Idiosincrático	Error del Modelo asociado a la varianza de $\mathbf{b}$	Error del modelo asociado a $\mathbf{S}$
Costa Urbana	0.38%	94.10%	5.52%
Costa Rural	0.22%	82.69%	17.09%
Sierra Urbana	0.24%	89.90%	9.86%
Sierra Rural	0.04%	79.16%	20.80%
Selva Urbana	0.54%	87.93%	11.53%
Selva Rural	0.12%	82.16%	17.72%
Lima Metropolitana	0.34%	93.23%	6.43%

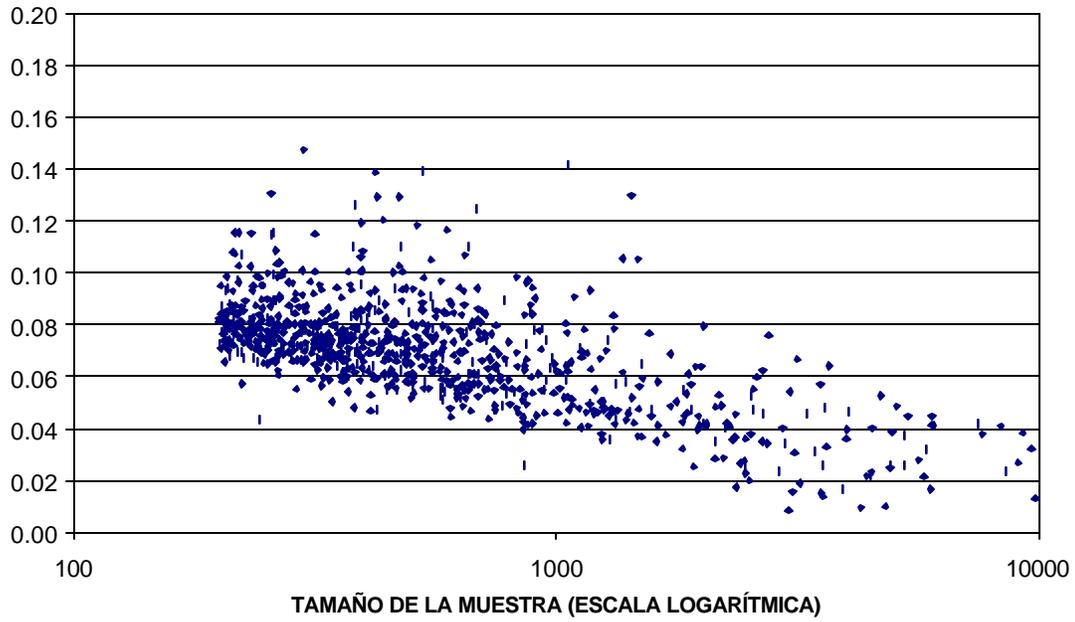
### **Estimaciones a niveles de agregación geográfica distrital y provincial**

A lo largo del documento se ha insistido en el problema de la falta de precisión de los indicadores de pobreza a niveles reducidos de agregación geográfica. Al respecto y luego de estimar los errores estándar de acuerdo a la metodología explicada en la sección 3.1.2., cabe resaltar que el ejercicio desarrollado confirma lo planteado por los documentos de Hentschel et al.(2000). y Elbers et al. (2000): parecería existir una relación inversa entre el tamaño poblacional del agregado geográfico de interés y la precisión con la cual es posible estimar el indicador de pobreza asociado a él.

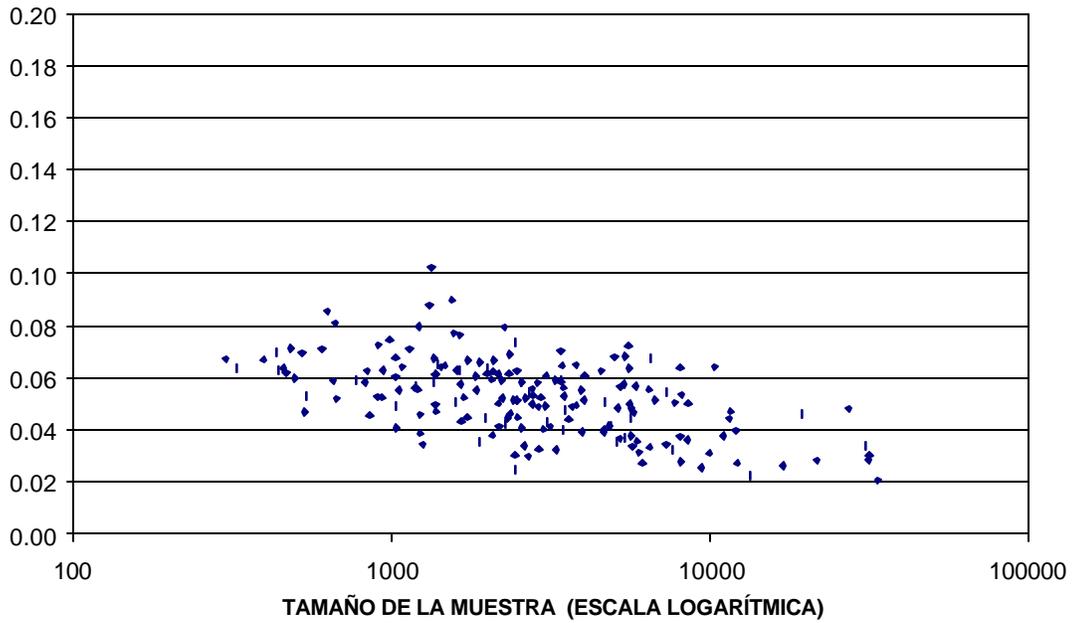
Los Gráficos 1 y 2 muestran esta relación inversa, obtenida en el marco del ejercicio desarrollado. Al respecto, cabe resaltar que los gráficos mostrados refieren a distritos para los cuales se contó con 200 o más reportes individuales. El Gráfico 1 muestra los resultados a nivel de distrito, mientras que el Gráfico 2, muestra los resultados a nivel provincial. De estos gráficos se confirma que a menor el tamaño de la muestra mayor el error estándar de la tasa de pobreza. Esto sugiere que a niveles reducidos de agregación geográfica la utilización de mapas de pobreza como instrumentos de focalización podría estar asociado a una menor precisión haciendo evidente la necesidad de medir los errores estándar para poder tener una idea de la magnitud del posible error a la hora de focalizar.

Otro aspecto, que se desprende de comparar el Gráfico 1 y el Gráfico 2, es que la variancia del error estándar es menor para el caso del ejercicio a nivel provincial (ver reducción en escala vertical) respecto del correspondiente a nivel distrital. Sin embargo, para poder evaluar si esta diferencia es significativa, es necesario complementar este análisis, formalizándolo a través de alguna prueba estadística.

**GRÁFICO 1**  
**ERRORES ESTÁNDAR DE LA TASA DE POBREZA**  
**(distritos con más de 200 reportes individuales)**

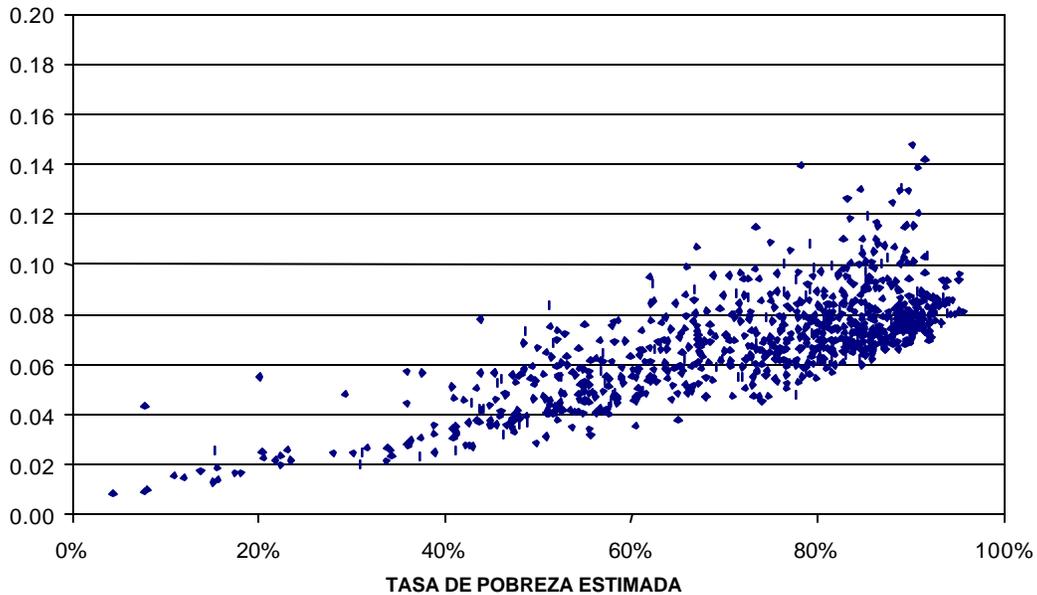


**GRÁFICO 2**  
**ERRORES ESTÁNDAR DE LA TASA DE POBREZA**  
**(provincias con más de 200 observaciones individuales)**

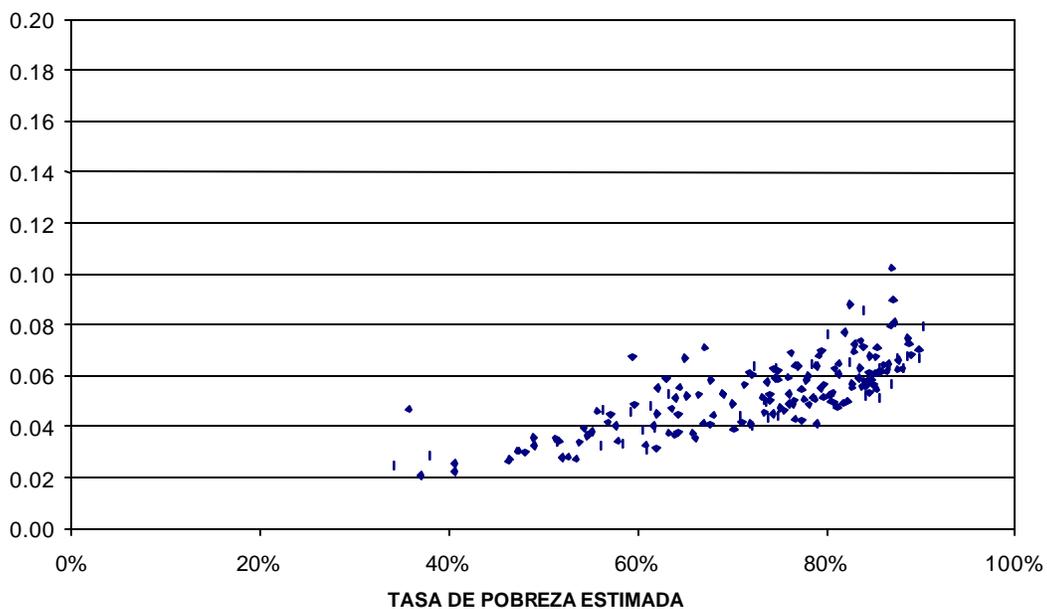


Asimismo, en el marco del ejercicio desarrollado, se observó una relación directa entre la tasa de pobreza estimada y el error estándar asociado a la misma (ver Gráficos 3 y 4). Es decir a medida que la tasa de pobreza se incrementa el error estándar tiende a ser mayor. Sería interesante explorar si esa tendencia se confirma para el caso de estimaciones que cuenten con el 100% de los reportes del censo.

**GRÁFICO 3**  
**ERRORES ESTÁNDAR DE LA TASA DE POBREZA**  
(distritos con más de 200 observaciones individuales)



**GRÁFICO 4**  
**ERRORES ESTÁNDAR DE LA TASA DE POBREZA**  
(provincias con más de 200 observaciones individuales)



En resumen, el ejercicio desarrollado muestra que los errores de predicción pueden ser significativos, especialmente cuando se reduce el nivel de agregación geográfica. Este hecho alerta sobre la importancia de la estimación del error de predicción de la tasa de pobreza y el valor que este puede tener para poder determinar los intervalos de confianza de dicho indicador.

## 5 CONCLUSIONES Y RECOMENDACIONES METODOLÓGICAS

Los mapas de pobreza constituyen el instrumento de focalización geográfica por excelencia y son utilizados comúnmente en consideración a sus bajos costos de construcción, implementación y monitoreo (en términos relativos a los exigidos por instrumentos de focalización individual). Su utilidad práctica radica en establecer ordenamientos basados en los niveles de bienestar de conglomerados poblacionales (típicamente a nivel de agregación geográfica distrital). Al respecto, cabe resaltar que los mapas de pobreza NO tienen como objetivo, y por tanto no son metodológicamente aptos para, aproximar relaciones de causalidad entre las variables utilizadas en la estimación de los niveles de bienestar y estos últimos. Debido a ello no deben ser utilizados para el diseño de las acciones de combate a la pobreza, si no exclusivamente en la identificación de conglomerados de poblaciones pobres.

De acuerdo a estudios especializados, las ganancias en eficiencia de focalización derivadas de la utilización de mapas de pobreza son significativas únicamente en caso de producir ordenamientos a niveles reducidos de agregación geográfica, al interior de los cuales exista cierto nivel de homogeneidad. A estos niveles de agregación, sin embargo, no se dispone de información sobre niveles de consumo o ingreso que permita la medición adecuada de los niveles de bienestar de los hogares. En este sentido, el presente documento ha tenido como objetivo central presentar una metodología de construcción de mapas de pobreza que permita integrar información de fuentes complementarias; y a partir de la cual sea posible efectuar estimaciones del bienestar de los hogares a través de información de consumo familiar, así como conocer los límites de las mismas a través de la estimación de los intervalos de confianza de los estimadores.

En esta línea, se exploró a nivel conceptual la metodología propuesta por Hentschel et al. (1998, 2000) y complementada por Elbers et al. (2000) y se desarrolló un ejercicio práctico de la misma haciendo uso de una Encuesta Nacional de Hogares sobre condiciones de vida y pobreza (INEI, ENAHO 98 IV Trimestre) y el 5% del Censo de Población y Vivienda de 1993. Ambas tareas han permitido desarrollar un cuerpo metodológico que apunta a mejorar la precisión en la estimación de indicadores de bienestar y el conocimiento de los niveles de agregación geográfica para los cuales se mantiene la validez de los mismos.

A continuación se presentan las principales recomendaciones metodológicas derivadas del trabajo de investigación en el que se sustenta este documento.

Para efectos de claridad expositiva, las recomendaciones metodológicas presentadas a continuación se estructuran de la misma manera que el marco metodológico desarrollado en la Parte 3 del documento:

- Indicadores para medir el bienestar de los hogares.
- Definición de la unidad de análisis y compatibilización de las fuentes de información utilizadas (censo y encuesta).
- Estimación de parámetros utilizados para estimar el indicador de bienestar.

- Incorporación del diseño muestral de la encuesta.
- Exploración de formas funcionales.
- Criterios de selección del modelo con mejor ajuste.
- Interpolación en la base de hogares censales de los parámetros estimados.
- Agregación del indicador de bienestar estimado por hogar a niveles diferenciados de agregación geográfica y estimación del intervalo de confianza correspondiente.

## **A Indicadores de Bienestar de los Hogares en la Construcción del Mapa de Pobreza**

1. Se recomienda utilizar indicadores de bienestar basados en el gasto per cápita del hogar - como indicadores de incidencia (o tasa de pobreza), brecha y severidad - en tanto su construcción tiene como correlato teórico la función de utilidad indirecta, con amplio sustento en la teoría económica.
2. La construcción de un mapa de pobreza basado en la estimación de un indicador de incidencia o tasa de pobreza requiere a nivel censal información de niveles de gastos de los hogares así como de las características socioeconómicas y demográficas de los mismos. Si bien es cierto, este tipo de información se puede recolectar a niveles geográficos muy reducidos, como un departamento o provincia, para hacer esta estimación, esta información sería muy costosa de obtener para todo el país. Así pues, para construir un mapa de pobreza a nivel nacional se requerirá de la utilización complementaria de por lo menos dos fuentes de información. Por un lado, una encuesta de hogares que incluya información sobre niveles de gasto de los hogares y otras características socio económicas y demográficas de los mismos. Por otro lado, un censo de población y vivienda que permita acceder a información sobre tales características a nivel de todos los hogares del país.

La metodología propuesta en el documento puede ser replicada para poblaciones más reducidas, como por ejemplo, en caso de contar con un censo departamental, es posible construir un mapa de pobreza distrital del departamento siempre que se garantice la representatividad a nivel departamental de la encuesta de hogares utilizada.

3. Con el objetivo de efectuar las estimaciones de un indicador de incidencia de la pobreza a niveles diferenciados de agregación geográfica, la estrategia global recomendada es la siguiente. En primer lugar, se estima una relación funcional entre el nivel de *gasto per cápita* de los hogares, reportado en las encuestas de hogares sobre niveles de vida y pobreza, y un conjunto de variables reportadas tanto en la encuesta de hogares como en el censo de población y vivienda utilizados. Con base en los parámetros que se obtienen de esta estimación se procede a efectuar la interpolación de los niveles de gasto de los hogares censales. Una vez realizada la interpolación se obtienen estimaciones de la tasa de pobreza a niveles de agregación geográfica diferentes, obteniendo los respectivos errores de predicción. Adicionalmente, y dado que los indicadores de tasa de pobreza que interesan en la elaboración del mapa son indicadores a nivel de individuos, se incorpora como

ponderador el tamaño del hogar en las estimaciones de tasa de pobreza y error de predicción.

## **B Definición de la Unidad de Análisis y Compatibilización de Fuentes de Información**

4. Es indispensable garantizar la compatibilidad de la información que será conjuntamente utilizada en las estimaciones del indicador de bienestar. En particular, la definición y compatibilización de la unidad de análisis constituye el elemento central en torno al cual se debe efectuar la compatibilización de la información contenida en ambas bases.
5. Se recomienda la utilización del hogar como unidad de análisis, en tanto el tema en cuestión es la Pobreza y la información de gasto recopilada en las encuestas de hogares está a nivel del hogar..
6. Es necesario evaluar las posibles definiciones de hogar (es decir, las posibles alternativas que definen a un individuo como miembro de un hogar) con las que es posible trabajar en el proceso de estimación del indicador de bienestar. Al respecto, es indispensable tomar en consideración dos criterios para elegir la definición de hogar que se utilizará en las estimaciones:
  - a. La definición de hogar debe ser conceptualmente compatible con el diseño muestral de la encuesta, en particular, con la utilización de pesos muestrales. Los pesos muestrales reflejan la probabilidad de ser encuestado de un hogar y, en consecuencia, la probabilidad de ser encuestado de cada uno de sus miembros. Al respecto, la caracterización de los miembros del hogar debe garantizar que cada uno de los individuos de la población pertenece a un solo hogar, y evitar por tanto la duplicación de pesos muestrales que sería conceptualmente inconsistente, y empíricamente contradictoria con el tamaño poblacional efectivo.
  - b. Por otro lado, la definición de la unidad de análisis debe permitir establecer con la mayor precisión posible el indicador de bienestar que será atribuido a los individuos que componen el hogar. Este aspecto es particularmente importante en el marco de esta metodología en tanto el indicador de bienestar utilizado se basa en el gasto per cápita del hogar, el cual es medido como la porción del gasto total del hogar que le corresponde a cada uno de los miembros que lo integran, parte que se asume equitativamente distribuida al interior del hogar.
7. En caso de efectuar modificaciones en la definición de hogar es indispensable redefinir los rubros de gasto per cápita atribuibles a los miembros de los hogares incluidos en la encuesta; así como recalcular la línea de pobreza utilizada (como se indicó en el documento, las estimaciones de pobreza pueden ser particularmente sensibles a estas redefiniciones; en particular, las referidas a la línea).

Se recomienda no efectuar ajustes sobre los pesos muestrales, en tanto la modificación del diseño muestral puede tener implicancias importantes en las estimaciones efectuadas y el nivel de confiabilidad de las mismas.

8. Asimismo, es posible que sea necesario redefinir las variables disponibles en las bases de datos de la encuesta y el censo en función de la definición de hogar por la que se haya optado (por ejemplo, el tamaño del hogar).
9. Adicionalmente a los cambios derivados de la redefinición de la unidad de análisis, es necesario compatibilizar las variables incluidas en ambas bases de datos; en particular, las categorías y la población sobre la cual se recogen los reportes en cada una de ellas. Por ejemplo, en el caso del nivel educativo de los miembros del hogar, una fuente de información puede desagregar la variable de nivel educativo en 5 categorías y la otra en 4 categorías; una puede recoger reportes de todos los miembros del hogar de 3 años y más, mientras la otra puede presentar reportes exclusivamente de miembros mayores de 6 años.

## **C Estimación de los Parámetros utilizados para estimar el Indicador de Bienestar**

### **Incorporación del Diseño Muestral**

10. Se considera necesaria la incorporación del diseño muestral en las estimaciones de los parámetros a interpolar en la base censal, en la medida que permite obtener estimadores insesgados y conocer los errores estándar correctos de tales estimadores. Tal como se señala en el documento, la incorporación parcial del diseño muestral (i.e. incluir solo pesos muestrales y no la especificación de estratos y unidades primarias de muestreo) puede llevar a subestimaciones importantes de los errores estándar de los estimadores, específicamente en muestras estratificadas como es el caso de las encuestas de hogares. En consecuencia el no incluir el diseño muestral puede llevar a incluir en el modelo variables (ya que sus  $t$ -estadísticos serían más grandes) que no contribuyen de manera importante a explicar la varianza total del gasto per cápita o la condición de pobreza. Más aún, también existe un factor de eficiencia, en el sentido que los términos del error en la regresión estarían correlacionados a lo largo de las observaciones por lo que un MICO ordinario no sería eficiente.

### **Exploración de Formas Funcionales**

11. La exploración de formas funcionales alternativas que mejoren la capacidad predictiva del modelo constituye uno de los aspectos centrales para mejorar la precisión de las estimaciones censales. Como se muestra en el documento, un componente importante del error de predicción de la tasa de pobreza estimada sobre los hogares censales se deriva de la forma funcional utilizada en la estimación de los parámetros.
12. Si bien el presente documento ha avanzado en explorar las posibilidades del modelo probit y el lineal (logarítmico), introduciendo interacciones entre variables para incorporar no linealidades en el gasto o en la probabilidad de ser pobre, es necesario reconocer que la selección del modelo con mejor ajuste dependerá de la base de datos utilizada, es decir que NO es posible definir a priori la superioridad de uno u otro modelo.

### **Criterios de Selección del Modelo con Mejor Ajuste**

13. Los criterios de selección del modelo con mejor ajuste propuestos en este documento son cuatro: (a) la diferencia entre la tasa de pobreza obtenida a partir de los reportes de la encuesta y la obtenida a partir de las predicciones de cada uno de los modelos; (b) la tasa de acierto de cada uno de los modelos; (c) el test de significancia de parámetros; y (d) el error estándar de las tasas de pobreza estimadas a partir de cada uno de los modelos. Adicionalmente, se recomienda explorar otros criterios que permitan evaluar la superioridad del modelo a utilizar en la interpolación. Un criterio adicional a explorar es el de validación de la predicción fuera de la muestra (i.e. efectuar la estimación de los modelos con mejor ajuste sobre reportes de otra encuesta de hogares).
14. En tanto la unidad de análisis es el hogar, se considera necesario evaluar los criterios de selección entre modelos haciendo uso de los factores de expansión de hogares, y no el correspondiente a la expansión de individuos.

Cabe recordar que tres de los cuatro criterios propuestos se basan en la agregación de las predicciones por hogar, de manera que la expansión a nivel de hogares permite prevenir de potenciales sesgos inducidos por diferencias en el tamaño de los hogares sobre los cuales se efectúan las estimaciones.

15. La evaluación de la diferencia entre la tasa de pobreza obtenida a partir de los reportes de la encuesta y la obtenida a partir de las predicciones de cada modelo debe considerar el intervalo de confianza de la primera. En caso de obtener diferencias que no excedan el intervalo de confianza referido, este criterio no debe ser tomado en cuenta en la selección del modelo, porque las diferencias observadas no serían estadísticamente significativas.

Al respecto, es necesario prestar especial atención sobre los sesgos que se pueden generar al construir el indicador de incidencia o tasa de pobreza. Como se indicó en la Parte 3 del documento, la predicción del nivel de consumo per cápita de los hogares, a través de la estimación lineal, produce estimadores sesgados de tasa de pobreza si estos son calculados mediante el conteo de hogares en situación de pobreza, e insesgados si se efectúan con base en la probabilidad de que un hogar sea pobre (dado su nivel de gasto estimado). Esto se debe a que la predicción del nivel de gasto de un hogar no permite tener certeza sobre su condición de pobreza.

16. La comparación de tasas de acierto constituye un criterio de evaluación de la capacidad predictiva del modelo más preciso que el anterior, en tanto evalúa la capacidad del modelo de predecir el nivel de bienestar de los hogares individualmente. En este caso, se compara el nivel de bienestar reportado y el predicho de cada uno de los hogares, uno a uno, y se pondera de acuerdo al peso muestral correspondiente.

Cabe enfatizar el análisis de los errores de predicción denominados de filtración (i.e. hogares no pobres predichos como pobres) y subcobertura (i.e. hogares pobres

predichos como no pobres) que puede ser derivado de este criterio. Dependerá del interés particular de la institución que hace uso del mapa, la priorización o no de modelos que minimicen alguno de estos tipos de error.

17. El test de significancia de parámetros recomendado es el test de Wald, que permite la comparación de modelos anidados. Al respecto, cabe enfatizar la pérdida de grados de libertad derivada de la introducción del diseño muestral de la encuesta; la misma que puede limitar el número de parámetros cuya significancia conjunta pretenda ser evaluada.
18. El cuarto criterio de evaluación propuesto es el referido a la magnitud de los errores estándar de las estimaciones de tasa de pobreza obtenidas en base a las predicciones de cada uno de los modelos, y permite estimar el intervalo de confianza de las estimaciones efectuadas. Dado que el objetivo final es predecir la tasa de pobreza a distintos niveles de agregación geográfica, es preferible un modelo que genere una predicción más precisa a otro que, estimando la misma tasa de pobreza, lo haga con un error estándar mayor.

#### **D Interpolación**

19. En referencia a la base de datos censal, es importante considerar efectuar imputaciones por la ausencia de reportes de ciertas variables. Por ejemplo, si un hogar reporta todas sus características excepto una, es recomendable imputar este dato y estimar la probabilidad de ser pobre de ese hogar. Para ello, puede hacerse uso de la media o mediana del centro poblado al que pertenece el hogar que requiere la imputación. Al respecto, se recomienda utilizar la mediana del centro poblado, u otro indicador adecuado, para evitar que la imputación esté afectada por valores extremos. Seguidamente, se recomienda evaluar que las imputaciones efectuadas no generen sesgos sistemáticos.

En caso de verificar que las imputaciones generan sesgos sistemáticos, es necesario identificar las observaciones (hogares) que, habiendo sido imputadas, son responsables de tales sesgos y eliminarlas de la base sobre la cual se pretende efectuar la interpolación.

20. La base censal utilizada en la interpolación debe contar exclusivamente con hogares que presenten reportes que no excedan los rangos de variación de los reportes de la encuesta para las variables pertinentes. Esto es así, debido a que no podemos pretender inferir comportamientos que se escapen del rango observado en la encuesta que sirvió para modelar la relación entre las características de los hogares y el indicador de bienestar.
21. En caso de no ser posible efectuar la interpolación de algunas observaciones (debido a que no fue posible efectuar imputaciones, o que los rangos de variación de las variables pertinentes excedieron los observados en la encuesta), es indispensable tomar nota del tamaño de los hogares perdidos con el fin de generar un factor de ajuste a nivel de centros poblados que permita recuperar el tamaño del hogar perdido. De esta manera, se garantiza la estimación de la tasa de pobreza sobre el tamaño poblacional efectivo del agregado geográfico de interés (centro poblado, distrito, provincia o departamento).

22. La interpolación de los parámetros estimados por el modelo seleccionado permite obtener una estimación del nivel de bienestar a nivel de cada uno de los hogares de la base censal.

### **E Agregación de Indicadores de Bienestar**

23. Una vez obtenidas las estimaciones del nivel de bienestar de los hogares censales, es necesario efectuar la agregación de estas estimaciones a diferentes niveles geográficos (centro poblado, distrito, provincia o departamento). Cabe resaltar que, a diferencia de la etapa de selección del modelo con mejor ajuste, en esta etapa se considera fundamental el sesgo derivado de la introducción del tamaño del hogar en la estimación de la tasa de pobreza para hallar la tasa a nivel individual.
24. El proceso de estimación del indicador de pobreza involucra dos fuentes de error (idiosincrático y del modelo) que limitan la confiabilidad del estimador, y con ello, su utilidad práctica. Por un lado, el error del modelo depende del ajuste del modelo seleccionado, y se deriva por tanto del diseño muestral y la forma funcional utilizada. Las sugerencias esbozadas en el punto C pretenden reducir este primer componente. Por otro lado, el error idiosincrático crece conforme se reduce el tamaño poblacional de la localidad en cuestión, por lo que limita la capacidad de desagregación geográfica de las estimaciones.
25. Debido a que el error de predicción aparentemente crece conforme se reduce el tamaño poblacional de la localidad para la cual se efectúan las estimaciones, es necesario establecer límites permisibles de error, con el fin de asegurar la validez de las estimaciones de pobreza sobre la base de las cuales se construye el mapa. El ejercicio realizado sugiere que solo en el caso de centros poblados muy grandes se lograría una estimación confiable, por lo que, en general, el mapa no deberá desagregarse más allá del nivel distrital.

## 6 BIBLIOGRAFÍA

Alderman, H; Babita, M, Lanjouw, J., Lanjouw (2000) "Is Census Income an Adequate Measure of Household Welfare?: Combining Census and Survey Data to Construct a Poverty Map of South Africa". Mimeo.

Bernedo, J. (1999) "La Selección de Pobres y el Mapa PROFINES. Problemas Y Alternativas". Cooperación Técnica Alemana / Ministerio de la Presidencia PROFINES. Lima.

BCRP (1986) "Mapa de Pobreza del Perú 1981". Subgerencia de Investigación Económica Banco Central de Reserva del Perú. Lima.

BCRP (1982) "Mapa de Pobreza del Perú". Reseña Económica: 45-58. Lima.

Baker, J. y M. Grosh (1994) "Proxy means test for targeting social programs". LSMS Working Papers 118. World Bank

Bigman D. and H. Fofack (2000) "Geographical Targeting for Poverty Alleviation. Methodology and Applications". World Bank Regional and Sectoral Studies.

Bigman, D. and H. Fofack (2000) "Geographical Targeting for Poverty Alleviation: An Introduction to the Special Issue". The World Bank Economic Review 14 (1):129-145.

Bigman, D., S. Dercon, D. Guillaume y M. Lambotte (2000) "Community Targeting for Poverty Reduction in Burkina Fasso". The World Bank Economic Review 14 (1):167-193.

Binswanger, H., S.R. Khandker y M.R. Rosenzweig (1993) "How Infraestructura and Financial Institutions Affect Agricultural Output and Investment in India". Journal of Development Economics 41(2): 337-366.

Chacaltana, J. (2000) "Más Allá de la Focalización: Performance e Impacto del Gasto (Focalizado) de Lucha contra la Pobreza". Lima: Consorcio de Investigación Económica. Mimeo.

Davidson, Russell y Mackinnon, James; (1993). Estimation and Inference in Econometrics. Oxford University press. Oxford.

Deaton, Angus; (1997). The Analysis of Household Surveys. A Microeconomic Approach to Development Policy. Johns Hopkins.

Elbers, C., P. Lanjow and J. Lanjow (2000) "Welfare in Villages and Towns: Micro-Level Estimation on Poverty and Inequality. World Bank.

Elías, Lidia (1992) "Estudios de Pobreza en el Perú: bibliografía comentada" Documentos de Trabajo No. 1. ADEC-ATC. Lima.

- Escobal, J. , Saavedra, J. y Torero, M. (1998) "Los activos de los pobres en el Perú". Documento de Trabajo 26. GRADE. Lima
- Escobal, J. y M. Torero (2000) "¿Cómo Enfrentar una Geografía Adversa?: el rol de los activos públicos y privados". Documento de Trabajo 29. GRADE. Lima
- Fofack, H. (2000) "Combining Light Monitoring Surveys with Integrated Surveys to Improve Targeting for Poverty Reduction: The Case of Ghana". The World Bank Economic Review 14 (1):196-211.
- FONCODES (1995) "El Mapa de la Inversión Social: FONCODES frente a la Pobreza 1991-1994". UNICEF. Lima.
- FONCODES (2000) "Mapa de Pobreza 2000. Principales Indicadores que definen el nivel de vida". FONCODES.
- Francke, P. y J. Guabloche (1995) "Focalización: identificando a los pobres según características socioeconómicas de los hogares". Documento de Trabajo 4. Gerencia de Estudios Económicos. Lima
- Foster, A. Y M.R. Rosenzweig (1995) "Learning by Doing from Others: Human Capital and Technical Change in Agriculture". Journal of Political Economy 103 (6): 1176-1209.
- Grosh, M. (1994) "Administering targeted social programs in Latin America: From Platitudes to Practice". World Bank.
- Haddad, L. y R. Kanbur (1992) "Intrahousehold Inequality and the Theory of Targeting". International Food Policy Research Institute (IFPRI) (reimpresión).
- Hentschel, J. and J. Lanjouw (1996) "Constructing an Indicator of Consumption for analysis of Poverty". LSMS Working Paper No. 124. World Bank.
- Hentschel, J., J. Lanjouw, P. Lanjouw and J. Poggi (2000) "Combining Census and Survey Data to Study Spatial Dimensions of Poverty: A Case Study of Ecuador". The World Bank Economic Review 14 (1):147-165.
- INEI (1992) "Censos Nacionales: IX de Población y IV de Vivienda 1993. Programa Censal". Lima.
- INEI (1995) "Proyecciones de la Población del Perú: 1995-2025". Boletín de Análisis Demográfico No. 34. Dirección Técnica de Demografía y Estudios Sociales. Lima.
- Instituto Cuánto (1995), "Retrato del familia peruana: Niveles de Vida, 1994". UNICEF. Lima.
- Medina (1999) Manual de Procedimientos para el Diseño y Selección de la Muestra de la ENAHO 98. Documentos Metodológicos. MECOVI-Perú. Lima.

Minot, N. (1998) "Generating Disaggregated Poverty Maps: An Application to Viet Nam". Markets and Structural Studies Division. International Food Policy Research Institute (IFPRI). Washington D.C.

Pferffermann, Danny; (1993). "The Role of Sampling Weights when Modeling Survey Data". International Statistical Review, Vol 61(2). pp.318-337.

Rao, J.N.K. y A.J. Scott (1981) The Analysis of Categorical Data from Complex Sample Surveys: Chi-Squared Tests for Goodness and Independence in Two-Way Tables. Journal of the American Statistical Association, Vol. 76, Issue 374: 221-230.

Ravallion, M. y M. Lipton (1993) "Poverty and policy". En: Working Papers, 1130, World Bank.

Ravallion, M. y Q. Woodon (1997) "Poor areas or only poor people?". Policy Research Working Paper 1798. World Bank.

Ravallion, M. y J. Jalan (1998) "Geographic Poverty Traps?". World Bank.

Robles, M. (1997) "Propuesta Metodológica para determinar el Ingreso y la Proporción de Hogares Pobres a Nivel Provincial y Distrital". INEI: Curso Taller Subregional sobre Indicadores Sociales de Pobreza, Mujer y Niñez. Lima.

Robles, M. (1999) "Indicadores Básicos para Focalizar el Gasto Social en Paraguay". BID / BM / CEPAL / MECOVI-Paraguay.

STATA 6.0 (1999) Reference Guide /User's Guide. Texas: Stata Press, College Station.

Saavedra, J. (1999) "Evaluación de la situación de la pobreza en el Perú y de los programas y políticas de lucha contra la pobreza". PNUD.

Sen, A. (1980) "Levels of Poverty: Policy and Changes". En: World Bank Staff Working Paper 401, July 1980.

Sen, A. (1997) "On Income Inequality". Oxford: Clarendon Press.

Shack, N. (1996) "Asignación de Recursos: un ejemplo de Focalización en el Perú: FONCODES". En: Información sobre Población y Pobreza para Programas Sociales, INEI. Lima.

Shack, N. y J. Díaz (2000) "Un Nuevo Instrumento de Focalización para la Asignación de Recursos Destinados a la Inversión Social en el Marco de la Lucha contra la Pobreza". Documento de Trabajo, Ministerio de Economía y Finanzas. Lima.

Schady, N. (1998) "Picking The Poor: Indicators for Geographic Targeting in Peru". Princeton University / World Bank.

Skinner, C.J. ; Holt, D; y Smith, T; (1989). Analysis of Complex Surveys. John Wiley & Sons, New York.

## 7 BASES DE DATOS

INEI Encuesta Nacional de Hogares 1998 del IV Trimestre:

- Base de Datos
- Manual de la Encuestadora y cédulas de la encuesta
- Documentos metodológicos :
  - Deflatación y Trimestralización de valores monetarios
  - Imputación de alquileres, último gasto por consumo, productos alimenticios, gasto del hogar, gasto en educación, gasto en salud, imputación de horas de trabajo, imputación de ingresos, de gastos en alimentos consumidos fuera del hogar
  - Trimestralización del ingreso y gasto monetarios (Sumaria)

INEI Muestra sistemática del 5% del Censo Nacional IX de Población y IV de Vivienda 1993:

- Base de datos
- Diccionario de Variables

## ANEXO 1

### *Apuntes Metodológicos del Ejercicio Práctico*

#### **A.1 Bases de datos**

Las bases de datos utilizadas fueron:

- (a) La Encuesta Nacional de Hogares 1998 del Cuarto Trimestre, realizada por el Instituto Nacional de Estadística e Informática, utilizada para estimar los parámetros que reflejan la relación entre determinadas características de cada hogar y el nivel de gasto reportado por el mismo. Esta base de datos presenta reportes de 6,952 hogares a nivel nacional, permitiendo representatividad de las estimaciones de pobreza a nivel de dominios (Costa Urbana, Costa Rural, Sierra Urbana, Sierra Rural, Selva Urbana, Selva Rural, Lima Metropolitana).
- (b) Una muestra sistemática (a nivel de departamento) del 5% del Censo Nacional IX de Población y IV de Vivienda 1993, la cual contiene reportes de 238,248 hogares a nivel nacional. Esta base de datos sirvió para efectuar la interpolación de los niveles de gasto per cápita de los hogares de la muestra y con ello el cálculo de las tasas de pobreza a distintos niveles de agregación (centros poblados, distritos, provincias, departamentos y dominios)

#### **A.2. Ajustes realizados**

En la presente sección se presenta el detalle de los principales ajustes realizados a las bases de datos (ENAHO y Censo). Cabe enfatizar que, además de ajustes específicos a las variables y unidades de análisis definidos en ambas bases de datos, se han efectuado modificaciones a la metodología de cálculo de la tasa de pobreza y pobreza extrema, siguiendo el método de Línea de Pobreza también utilizado por el INEI.

1. En una primera fase, se efectuaron ajustes en la base de datos de la Encuesta Nacional de Hogares (ENAHO) con el fin de compatibilizar sus reportes con los de la base censal. Estos ajustes apuntaron fundamentalmente a la definición de las unidades de análisis (hogares) y de las variables en función de las cuales se efectuarían las estimaciones y posteriormente la interpolación.
2. En una segunda fase, se realizaron algunos ajustes a la estimación de los niveles de gasto per cápita de los hogares encuestados, así como a la estimación de las tasas de pobreza reportadas en la encuesta (en base a cálculos del Instituto Nacional de Estadística e Informática - INEI). En base a los ajustes efectuados en los dos pasos anteriores, se armó la base de datos de hogares, en función de la cual se estimaron los parámetros de la función de gasto, los cuales serían interpolados en la base de datos censal.
3. La interpolación de los niveles de gasto de los hogares censales requirió tres ajustes adicionales en la base de datos del censo, (a) la homogeneización de los rangos de variación de los variables utilizadas para estimar el gasto per cápita respecto de los observados en la base de la encuesta de hogares, (b) la recuperación del mayor número de hogares posible en base a las medianas de las variables por centro

poblado o, en su defecto, por distrito, y (c) la aplicación de un factor de ajuste por las observaciones pedidas en (a) y (b).

Una vez efectuada la interpolación de los niveles de gasto per cápita correspondientes a los hogares censales se calcularon las tasas de pobreza por distrito, por provincia, por departamento y por dominio.

### **A.2.1. Ajustes en la base de datos de la Encuesta**

Como se mencionó líneas antes, la compatibilización de los reportes del censo y la encuesta exigió algunos ajustes importantes, los cuales son explicados a continuación.

#### A.2.1.1 Compatibilidad de las unidades de análisis

Como es conocido, tanto los reportes del censo como los de la ENAHO son tomados a tres niveles de agregación: vivienda, hogar e individuo. El problema de compatibilidad de unidades de análisis aparece en el caso del hogar; es decir, la definición de hogar censal no coincide exactamente con ninguna de las definiciones de hogar que aparecen de manera explícita en la encuesta.

En tanto el hogar constituye la unidad de análisis para la elaboración de mapas de pobreza, requerimos una definición de hogar en la encuesta con la cual trabajar las regresiones e interpolar los niveles de gasto en la base de hogares censales.

Según la cédula censal, el *hogar* está conformado por las personas que se encuentran presentes en la vivienda a partir de la noche anterior al día del censo<sup>38</sup>. En tanto esta definición no permite distinguir a los individuos que son miembros del hogar familiar de los que no lo son, se trató de calzar la definición de hogar que aparece en la base de datos de ENAHO a la que aparece en la base censal. Con este objetivo se analizaron las dos alternativas que aparecen en el Manual de la Encuestadora (ENAHO 1998-IV):

Hogar Familiar – Conjunto de personas, sean o no parientes (padres, hijos solteros, hijos casados, hermanos, tíos, etc.), que ocupan en su totalidad o en parte una vivienda, comparten las comidas principales y atienden en común otras necesidades vitales. Se incluye a personas a quienes el jefe de hogar considera que son miembros del hogar por razones de afecto (ahijados, compadres, padrinos, etc.).

Hogar ENAHO (definición original) – Constituido por los miembros del hogar familiar, los trabajadores del hogar con cama adentro (reciban o no pago por sus servicios), los integrantes de una pensión familiar que tienen como máximo 9 participantes, y las personas que no son miembros del hogar familiar, pero que se encuentran presentes en el hogar los últimos 30 días.

---

38 Cabe indicar que, en estricto, el INEI define hogar censal como “.. el conjunto de personas, sean parientes o no, que ocupan en su totalidad o en parte una vivienda, comparten las comidas principales y atienden en común otras necesidades vitales básicas.” (INEI, 1992: 22). Sin embargo, los reportes por hogar que Figuran en la base de datos del censo son tomados en base a la siguiente pregunta: “En este hogar, incluyendo a los recién nacidos y ancianos, ¿cuántas personas pasaron aquí la noche anterior al día del censo?” (Cuarta Sección de la cédula censal “Personas que conforman el hogar”). Debido a que la base del censo está construida en base a la pregunta que aparece en la cédula, no es posible distinguir entre los individuos que pertenecen al hogar y aquellos que se encuentran de paso.

Si bien la unidad de observación de la encuesta es el hogar ENAHO (es decir, se encuesta a los integrantes del hogar ENAHO, no únicamente a los miembros del hogar familiar), ésta no es compatible conceptualmente con la expansión de los reportes a población total<sup>39</sup>. Por ello, se reformuló la definición de hogar ENAHO de manera tal que (a) mantenga la característica de incluir a las personas que residen habitualmente en el hogar, (b) sea compatible con la aplicación de un factor de expansión que siga reflejando la probabilidad de ser encuestado del hogar<sup>40</sup>, y (c) pueda ser utilizada para tabular datos a nivel de población total<sup>41</sup>.

Hogar ENAHO (ajustado) – Constituido por los residentes habituales del hogar. Incluye a los miembros del hogar familiar que residen habitualmente en el hogar (se excluye a los que se encuentran ausentes 30 días o más), y a los que no son miembros del hogar familiar pero residen habitualmente en él (están presentes hace 30 días o más)<sup>42</sup>.

Así, tomando en consideración las definiciones de hogar familiar y hogar ENAHO, la expansión de los reportes a la población total da los siguientes resultados<sup>43</sup>:

Hogar ENAHO: 24'999,675 de habitantes

Hogar familiar: 25'145,606 de habitantes

Como es obvio, la expansión al total de hogares para ambas definiciones es la misma, y asciende a 5'396,580 hogares a nivel nacional.

### *Elección de la unidad de análisis de la encuesta*

La unidad de análisis elegida para el procesamiento de los reportes de la encuesta de hogares fue la correspondiente a *hogar familiar*. Las ventajas y desventajas más importantes de cada definición se presentan a continuación.

---

39 De hecho, la expansión a población total se hace usualmente en base a la definición del hogar familiar (documentos del INEI sobre estudios de pobreza en base a la ENAHO 98-4).

40 Cabe enfatizar que estas definiciones de hogar familiar y hogar ENAHO modifican el factor de expansión aplicable a los individuos encuestados, sin embargo, el hogar mantiene el mismo factor de expansión en ambos casos (debido a que está determinado por el conglomerado al que pertenece el hogar).

41 Al respecto, es importante tomar en consideración que el miembro de un hogar no puede ser miembro de otro hogar (lo que sucedería con los no miembros del hogar familiar que residen habitualmente en él, ya que estos individuos pertenecerían a este hogar como residentes habituales pero también pertenecerían a otro hogar familiar (esta vez como miembros ausentes del hogar familiar), bajo la definición original de hogar ENAHO).

42 En adelante, se hará referencia a esta definición como hogar ENAHO (y a la definición del Manual de la Encuestadora como definición de hogar ENAHO original).

43 Según proyecciones de población del INEI entre el 98 y el 99 habría un incremento de 24.8 a 25.2 millones de habitantes a nivel nacional, por lo que ambos resultados son aceptables. Cabe notar que, en caso de utilizar la definición original de hogar ENAHO (que incluye a miembros del hogar familiar ausentes), la expansión a población total ascendería a 25'584,359 de habitantes a nivel nacional.

Tipo de Hogar	Ventajas	Desventajas
<b>Hogar Familiar</b>	El gasto per cápita estimado es más cercano al que efectivamente beneficia a cada miembro de este hogar.	Proporciona información sobre las características de miembros del hogar familiar que se encuentran ausentes, de las cuales no se dispone en la base de datos censal. No proporciona información sobre las características de residentes habituales no miembros del hogar familiar ni de personas en tránsito (las cuales aparecerán en la fase de interpolación, por estar incluidas en la definición de hogar censal).
<b>Hogar ENAHO</b>	La definición es más cercana a la de hogar censal (residentes habituales del hogar, sean o no miembros del hogar familiar). En comparación con el hogar censal, la única diferencia estaría dada por la exclusión de las personas en tránsito (las que sí están incluidas en el hogar censal).	Tiende a sobre estimar el gasto per cápita correspondiente a residentes habituales que no pertenecen al hogar familiar, en especial en lo referido a rubros de gasto reportados a nivel de hogar (por el jefe de hogar familiar). No proporciona información sobre las características de personas en tránsito (las cuales aparecerán en la fase de interpolación, por estar incluidas en la definición de hogar censal).

Al respecto, cabe enfatizar que ambas alternativas son consideradas conceptualmente válidas, en tanto:

- (a) Ambas son compatibles con la aplicación del factor de expansión de hogares que refleja la probabilidad de ser encuestado. Este aspecto es de especial importancia en el marco de la metodología de interpolación desarrollada para la elaboración del mapa de pobreza.
- (b) En ambos casos, la expansión a población total, derivada de la aplicación de los referidos factores, es compatible con las proyecciones poblacionales del INEI para el año 1998.

Finalmente, cabe indicar que no se observan diferencias significativas en el tamaño promedio del hogar en caso de aplicar de manera alternativa una u otra definición.

Dominios	Tamaño Promedio del Hogar	
	Hogar ENAHO	Hogar Familiar
Costa Urbana	5.72	5.68
Costa Rural	5.97	6.06
Sierra Urbana	5.47	5.53
Sierra Rural	5.84	5.93
Selva Urbana	5.76	5.81
Selva Rural	6.66	6.80
Lima Metropolitana	5.58	5.55
Total	5.76	5.79

**Fuente: ENAHO 98-4**

### **A.2.1.2 Compatibilidad de las variables: homogeneización en la definición de las variables originales**

Se realizaron ajustes en la definición de las variables de las bases de datos originales que servirían de base para la generación de las variables utilizadas en la interpolación. Los principales ajustes se hicieron a nivel de rangos de edad para los cuales eran válidos los reportes de la encuesta y el censo. Por ejemplo, en relación a los datos sobre ocupación y actividad económica de los miembros del hogar, la encuesta presenta reportes solo para individuos de 14 años y más, mientras el censo dispone de información para individuos de 6 años y más; en consecuencia, las variables del censo a utilizar en la interpolación fueron generadas únicamente para individuos de 14 años y más.

Al final del presente anexo se encuentra un cuadro con las diferencias entre variables originales de la encuesta y del censo, así como los ajustes efectuados para homogeneizar las variables de ambas bases.

### **A.3. Ajustes metodológicos a las estimaciones de pobreza con base en la ENAHO 1998 del IV Trimestre**

Luego de evaluar a profundidad la metodología de cálculo de pobreza del INEI, se optó por realizar cambios tanto en la estimación del gasto per cápita como en la línea de pobreza<sup>44</sup>.

#### **A.3.1 Modificación de la variable gasto per cápita**

Se efectuó una revisión de la agregación de gastos por hogar en la base de datos de la encuesta. Al respecto, se recalculó el nivel de gasto por hogar familiar, a lo cual le siguió un re-cálculo del gasto per cápita por hogar familiar.

<sup>44</sup> Cabe recordar que estas modificaciones se realizaron tomando como base la definición de hogar familiar.

Re-cálculo del gasto per cápita del hogar familiar:

- a. En primer lugar, se diferenci6 el ‘reporte individual’ de miembros del hogar familiar del que corresponde a residentes habituales que no son miembros del hogar familiar (estos reportes a nivel de individuo se encuentran en los m6dulos de empleo, salud y educaci6n, principalmente).
- b. Adicionalmente, se diferenci6 el ‘reporte por hogar’ correspondiente a los gastos que benefician exclusivamente a miembros del hogar familiar de aquellos que benefician tambi6n a los no miembros que residen habitualmente en 6l (este tipo de reporte aparece en el sexto m6dulo, y es tomado 6nicamente al jefe de hogar o ama de casa, en representaci6n de los dem6s residentes del hogar).
- c. Se recalcul6 el gasto del hogar familiar sumando los ‘reportes individuales’ de miembros del hogar familiar y la porci6n de los ‘reportes por hogar’ que se consider6 beneficiaban 6nicamente a los miembros del hogar familiar. En el cuadro que sigue se presenta un listado de los rubros reportados en la encuesta.

**Definici6n de rubros de gasto del hogar familiar que benefician exclusivamente a sus miembros  
(‘reportes por hogar’ y ‘reportes individuales’)**

<b>Benefician exclusivamente a los miembros del hogar familiar</b>	<b>Benefician a miembros del hogar familiar y a no miembros que residen habitualmente en el hogar</b>
<ul style="list-style-type: none"> <li>▪ Gastos por alimentos consumidos fuera del hogar – Pag6</li> <li>▪ Ingresos y Gastos por alimentos consumidos fuera del hogar</li> <li>▪ Vestido y Calzado - Gasto</li> <li>▪ Vestido y Calzado-Autosumin., pago en especie</li> <li>▪ Vestido y Calzado-Donado, regalado, otro</li> <li>▪ Vestido y Calzado-No gasto, no sabe</li> <li>▪ Cuidado, Conservaci6n Salud - Gasto</li> <li>▪ Cuidados de Salud. Autosuministro - pago especie</li> <li>▪ Cuidados de Salud - donado, regalado, otros</li> <li>▪ Cuidados de Salud - No gast6, no sabe</li> <li>▪ Transportes y Comunicaciones - Gasto</li> <li>▪ Transportes - Autosuministro pago en especie</li> <li>▪ Transportes - Donado, regalado, otro</li> <li>▪ Transportes y Comunicac. - No gasto, no sabe</li> <li>▪ Esparcimiento, diversi6n - Gasto</li> <li>▪ Esparcimiento, Diversi6n - Pago en especies</li> <li>▪ Esparcimiento, Diversi6n - Donado, regalado.</li> <li>▪ Esparcimiento, Diversi6n - No gast6, no sabe</li> <li>▪ Otros bienes y servicios - gasto</li> <li>▪ Otros Bienes - Autosuministro, pago en especie</li> <li>▪ Otros Bienes y Servicios - Donado, regalado, otros</li> <li>▪ Otros Bienes y Servicios - No gast6, no sabe</li> <li>▪ Gastos por equipamiento del hogar</li> </ul>	<ul style="list-style-type: none"> <li>▪ Alimentos-Gasto</li> <li>▪ Alimentos-Autosuministro, pago en especie</li> <li>▪ Alimentos-donado, regalado, otro</li> <li>▪ Gastos por alimentos para consumir dentro del hogar</li> <li>▪ Ingresos y Gastos por alimentos para consumir en el hogar</li> <li>▪ Alquiler Vivienda, Combustible - Gasto</li> <li>▪ Alquiler Vivienda, Combustible. - pago en especie</li> <li>▪ Alquiler Vivienda, Combustible, Electricidad,...</li> <li>▪ Alquiler Vivienda - No gast6, no sabe</li> <li>▪ Muebles y Enseres, Mantenim. Vivienda – Gasto</li> <li>▪ Muebles - Autosuministro pago en especie</li> <li>▪ Muebles y enseres - Donado, regalado, otros.</li> <li>▪ Muebles y enseres - No gast6, no sabe</li> </ul>

- d. Finalmente, se dividió el gasto del hogar familiar (recalculado) entre el número de miembros del hogar familiar.

### A.3.2 Modificación de la línea de pobreza extrema y no extrema (canasta básica alimentaria)<sup>45</sup>

Con respecto a la estimación de las tasas de pobreza (extrema y no extrema), cabe señalar que, si bien se efectuaron ajustes, estos fueron hechos siguiendo los criterios delineados por el INEI para la elaboración de las líneas de pobreza extrema (canasta básica alimentaria) y de pobreza (canasta básica de consumo).

Ajustes metodológicos esto es solo para pobreza extrema o para pobreza también

El INEI establece un requerimiento calórico mínimo único a nivel nacional, el cual asciende a 2,318 calorías per cápita diarias, las cuales consideran la composición etárea y de género a nivel nacional, en base a información censal.

Al respecto, se decidió establecer nuevos requerimientos calóricos mínimos tomando en consideración las diferencias en la actual (1998) composición de edad y de género de los hogares de Costa, Sierra y Selva (nivel de agregación representativo en la encuesta y en base al cual el INEI estableció las canastas básicas alimentarias en 1997, las cuales se mantuvieron para la elaboración de la línea de pobreza extrema utilizada en la ENAHO 98-4).

Se tomaron en cuenta los siguientes requerimientos mínimos por edad y género:

<b>Edad y género</b>	<b>Requerimiento calórico mínimo</b>
Hombre mayor de 18 años	3,000 calorías diarias
Mujer mayor de 18 años	2,695 calorías diarias
De 12 a 17 años	2,200 calorías diarias
De 6 a 11 años	2,380 calorías diarias
De 3 a 5 años	1,830 calorías diarias
De 1 a 2 años	1,360 calorías diarias
Menor de 1 año	820 calorías diarias

De acuerdo a la composición etárea y de género por regiones, los requerimientos calóricos mínimos establecidos fueron los siguientes:

Costa	2456.4 calorías diarias
Sierra	2421.1 calorías diarias
Selva	2416.1 calorías diarias

<sup>45</sup> Los cálculos presentados en esta parte fueron efectuados sobre la base de datos original de la encuesta, es decir sobre la totalidad de hogares, sin excluir aquellos que presentaron reportes con extremos.

Estos requerimientos calóricos mínimos fueron valorizados a precios de cada dominio (de acuerdo a los niveles de precios utilizados por el INEI en sus cálculos de pobreza de 1998). Adicionalmente, fueron utilizados para establecer una nueva población de referencia (en función del equivalente monetario de los requerimientos mínimos en mención), en base a la cual se estimaron los nuevos coeficientes de Engel por dominio.

Cabe señalar que no se efectuaron modificaciones a la composición de la canasta básica alimentaria propuesta por el INEI, la cual fue establecida para el año 1997 y mantenida para los cálculos del año 1998.

De acuerdo a la valorización referida, las líneas de pobreza extrema por dominios fueron las siguientes:

<b>Valor Trimestral de la Canasta Básica Alimentaria per cápita, por dominios (a nuevos soles de noviembre de 1998)</b>		
<b>Dominios</b>	<b>Ajustada</b>	<b>INEI</b>
Costa Urbana	313.69	296.01
Costa Rural	301.70	284.70
Sierra Urbano	280.89	268.83
Sierra Rural	250.39	239.64
Selva Urbano	293.18	281.28
Selva Rural	284.67	273.12
Lima Metropolitana	367.03	346.35

Fuente: ENAHO 98-4

### ***A.3.3. Modificación de la línea de pobreza absoluta (canasta básica de consumo, que incluye consumo alimentario y no alimentario)***

El INEI valoriza la canasta básica de consumo en base al coeficiente de Engel por dominio, sin embargo, en este estudio se optó por utilizar el coeficiente de Engel de Lima para evitar subestimaciones de la tasa de pobreza derivadas del reducido nivel de gasto en rubros no alimentarios en zonas con mayores niveles de pobreza, especialmente en áreas rurales del país.

Las líneas de pobreza derivadas de este ajuste se señalan en el siguiente cuadro.

**Valor Trimestral de la Canasta Básica de Consumo  
per cápita, por dominios  
(a nuevos soles de noviembre de 1998)**

Dominios	Ajustado	INEI
Costa Urbana	598.02	498.66
Costa Rural	575.17	423.15
Sierra Urbano	535.49	493.95
Sierra Rural	477.35	326.70
Selva Urbano	558.92	481.86
Selva Rural	542.7	375.39
Lima Metropolitana	699.72	653.34

Fuente: ENAHO 98-4

#### A.3.4. Tasa de Pobreza considerando los ajustes efectuados

Producto de los ajustes señalados anteriormente, se estimó una tasa de pobreza a nivel nacional de 49.4%, mayor a la resultante de los cálculos del INEI que asciende a 37.3%. A continuación se observan las tasas de pobreza por dominios derivadas de nuestros cálculos.

Dominios	Tasa de Pobreza Extrema		Tasa de Pobreza Total	
	Ajustada	INEI	Ajustada	INEI
Costa Urbana	6.5%	5.1%	42.2%	27.9%
Costa Rural	22.0%	19.9%	67.7%	43.1%
Sierra Urbana	9.9%	8.6%	37.3%	32.8%
Sierra Rural	45.7%	42.9%	80.2%	61.9%
Selva Urbana	8.7%	6.7%	44.2%	33.5%
Selva Rural	28.7%	26.2%	79.2%	53.2%
Lima Metropolitana	2.6%	2.0%	24.6%	21.2%
Nacional	17.3%	15.6%	49.4%	37.3%

Fuente: ENAHO 98-4

Con base en estas estimaciones se estimaron los parámetros a interpolar en la base del censo.

#### A.4. Ajustes a la base de datos del Censo

A continuación se detallan los principales ajustes efectuados a la base de datos del censo<sup>46</sup>.

<sup>46</sup> Cabe recordar que la homogeneización en la definición de variables indicada en 1.2 es válida tanto para la base de la encuesta como para la base censal.

#### **A.4.1 Homogeneización de los rangos de las variables construidas**

Una vez construidas las bases de datos (de la encuesta y el censo) con base en las cuales se trabajaría la interpolación, se calcularon los rangos de variación de cada una de las variables pertinentes. La interpolación se hizo dentro de los intervalos en los que se encontraban las variables reportadas en la encuesta, pues la interpolación en el censo, en base a los parámetros resultantes de las regresiones, es válida sólo para tales intervalos.

Las variables pertinentes fueron *mienhog* (número de miembros del hogar : se excluyó a los hogares censales que tuvieran más de 18 miembros), *numhab* (número de habitaciones de la vivienda : se excluyó a los que tuvieran más de 13 habitaciones), *tothog* (número de hogares en la vivienda: se excluyó a los hogares que pertenecieran a una vivienda ocupada por 5 hogares o más), *ntrabhog* (número de trabajadores del hogar en el hogar censal : se excluyó a los que tuvieran más de 2 trabajadores del hogar), *phom0a13* (proporción de miembros de sexo masculino: se excluyó a hogares que tuvieran una proporción mayor a 80%), *pmuj0a13* (proporción de miembros de sexo femenino: se excluyó a hogares que tuvieran una proporción mayor al 75%), *npension* (número de pensionistas en la vivienda: se excluyó a hogares que reportaran más de 2 pensionistas), *edadjef* (edad del jefe de hogar: se excluyó a los hogares con jefes de hogar menores de 16 años). Como resultado, se excluyó a un 4.3% del total de hogares censales disponibles en la base de datos (10,166 de los 238,248 hogares censales originalmente disponibles).

#### **A.4.2 Reemplazo de no-reportes por medianas de los valores de las variables a nivel de centro poblado o distrito**

Uno de los problemas más importantes en el momento de la interpolación es el referido al porcentaje de *no-reportes* en la base censal, en tanto el *no-reporte* de una o más variables por parte de un hogar imposibilita la interpolación del nivel de gasto del mismo. La base de datos disponible presentaba este problema para el 33.5% de los hogares censales incluidos en ella (76,318 de los 228,082 hogares de la base censal no reportaban información para una o más de las variables utilizadas en la interpolación).

Con el objetivo de minimizar la pérdida de hogares censales se efectuó un primer reemplazo de los *no-reportes* por las medianas de las respectivas variables a nivel de centro poblado. Este reemplazo permitió reducir el porcentaje de hogares para los que no era posible interpolar niveles de gasto de 33.5% a 6.3% del total de hogares disponibles en la base censal. A pesar de esta importante reducción, el número de hogares perdidos seguía siendo significativa, por lo que se optó por reemplazar los *no-reportes* restantes con las medianas de las correspondientes variables a nivel de distrito. Con este reemplazo fue posible reducir el porcentaje de hogares perdidos al 0.1% de los hogares censales incluidos en la base.

Como producto de los dos ajustes mencionados, fue posible interpolar los niveles de gasto de 227,895 hogares censales.

### **A.4.3. Aplicación de factores de ajuste**

El cálculo de las tasas de pobreza a nivel de centros poblados, distritos y departamentos requirió de la formulación de un factor de ajuste que permitiera recuperar los hogares censales eliminados de la base antes o después de la interpolación (debido a que presentaban variables cuyos rangos excedían los de sus correspondientes en ENAHO (ver punto anterior), o a que no fue posible interpolar sus niveles de gasto):

$$\text{Factor de ajuste} = \frac{\text{pob del centro poblado (base original)}}{\text{pob del centro poblado para la cual fue posible interpolar niveles de gasto}}$$

## ANEXO 2

COMPARACIÓN DE VARIABLES DE LA ENAHO Y EL CENSO		
CENSO 93	ENAHO 98-4	AJUSTES EFECTUADOS
Unidades de observación (u.o.): Vivienda : la información es reportada únicamente por el hogar principal. Hogar : la información es reportada por cada hogar particular (principal o secundario). Individuo : salvo especificaciones de rangos de edad, alude a los reportes individuales efectuados por todos los miembros del hogar		
u.o. – vivienda  Tipo de vivienda (para viviendas particulares) (casa independiente, dpto en edificio, viv en quinta, viv en casa de vecindad (callejón, solar, corralón), choza o cabaña, vivienda improvisada, local no dest para la hab humana, otro)	u.o. – vivienda  Tipo de vivienda (casa independiente, dpto en edificio, viv en quinta, viv en casa de vecindad (callejón, solar, corralón), choza o cabaña, vivienda improvisada, local no destinado para la hab humana, otro)	Categorías idénticas  <b>Problema:</b> Los hogares secundarios en la ENAHO presentan missings (se encuesta solo al hogar principal). <b>Ajuste efectuado:</b> Se imputó el reporte del hogar principal a los hogares secundarios de ENAHO.
u.o. – vivienda  Propiedad de la vivienda (alquilada, propia-comprándola a plazos, propia – totalmente pagada, usada con autorización del propietario – sin pago, ocupada de hecho (invasión), otros)	u.o. - hogar  Propiedad de la vivienda (alquilada, propia-comprándola a plazos, propia – totalmente pagada, cedida por el centro de trabajo, cedida por otro hogar o institución, propia por invasión, otros)	<b>Problema:</b> 1. El Censo reporta a nivel de vivienda (datos del hogar principal imputados a los secundarios) y la ENAHO a nivel de hogar. <i>Agravante:</i> Se presentan diferencias importantes entre reportes de hogares que pertenecen a la misma vivienda: Solo 18 hogares de los 6,952 (6 viviendas con 2 hogares c/u y en 2 viviendas con 3 hogares c/u) reportan el mismo dato al interior de c/vivienda (de un total de viviendas de 6,839). <b>Ajuste efectuado:</b> 1. Se decidió mantener los reportes por hogar en ENAHO. 2. Las categorías fueron agrupadas de la siguiente manera: (a) alquilada, (b) propia (propia-comprándola a plazos, propia –

COMPARACIÓN DE VARIABLES DE LA ENAHO Y EL CENSO		
CENSO 93	ENAHO 98-4	AJUSTES EFECTUADOS
		totalmente pagada), (c) resto (incluye las categorías cedida por el centro de trabajo, cedida por otro hogar o institución, propia por invasión, usada con autorización del propietario – sin pago, ocupada de hecho (invasión), otros)
u.o. – vivienda  Paredes exteriores (ladrillo o bloque de cemento, piedra o sillar con cal o cemento, adobe o tapia, quincha (caña con barro), piedra con barro, madera, estera, otros)	u.o. – vivienda  Paredes exteriores (ladrillo o bloque de cemento, piedra o sillar con cal o cemento, adobe o tapia, quincha (caña con barro), piedra con barro, madera, estera, otros)	Categorías idénticas  <b>Problema:</b> Los hogares secundarios en la ENAHO presentan missings (se encuesta solo al hogar principal). <b>Ajuste efectuado:</b> Se imputó el reporte del hogar principal a los hogares secundarios de ENAHO.
u.o. - vivienda  Pisos (parquet o madera pulida, láminas asfálticas – vinílicos – similares, losetas – terrazos – similares, madera (entablados), cemento, tierra, otros)	u.o. - vivienda  Pisos (parquet o madera pulida, láminas asfálticas – vinílicos – similares, losetas – terrazos – similares, madera (entablados), cemento, tierra, otros)	Categorías idénticas  <b>Problema:</b> Los hogares secundarios en la ENAHO presentan missings (se encuesta solo al hogar principal). <b>Ajuste efectuado:</b> Se imputó el reporte del hogar principal a los hogares secundarios de ENAHO.
u.o. - vivienda  Abastecimiento de agua de la vivienda (red pública dentro de la vivienda, red pública fuera de la vivienda pero dentro del edificio, pilón de uso público, camión – cisterna u otro similar, pozo, río – acequia- manantial – similar, otro)	u.o. - hogar  Abastecimiento de agua en el hogar (red pública dentro de la vivienda, red pública fuera de la vivienda pero dentro del edificio, pilón de uso público, camión – cisterna u otro similar, pozo, río – acequia- manantial – similar, otro)	<b>Problema:</b> El Censo reporta a nivel de vivienda (datos del hogar principal imputados a los secundarios) y la ENAHO a nivel de hogar. <i>Atenuante:</i> No se presentan diferencias importantes entre reportes de hogares que pertenecen a la misma vivienda: Solo al interior de 10 viv se presentan diferencias. <b>Ajuste efectuado:</b> Se decidió mantener los reportes por hogar en ENAHO.

**COMPARACIÓN DE VARIABLES DE LA ENAHO Y EL CENSO**

CENSO 93	ENAHO 98-4	AJUSTES EFECTUADOS
<p>u.o. - vivienda</p> <p>La vivienda tiene servicio higiénico (water, excusado, letrina, etc.) conectado a (red pública dentro de la vivienda, red pública fuera de la vivienda pero dentro del edificio, pozo ciego o negro, sobre acequia o canal, no tiene)</p>	<p>u.o. - hogar</p> <p>La vivienda tiene servicio higiénico (water, excusado, letrina, etc.) conectado a (red pública dentro de la vivienda, red pública fuera de la vivienda pero dentro del edificio, pozo ciego o negro – letrina, sobre río - acequia - canal, no tiene)</p>	<p><b>Problema:</b> Censo reporta a nivel de vivienda (datos del hogar principal imputados a los secundarios) y la ENAHO a nivel de hogar. <i>Atenuante:</i> No se presentan diferencias importantes entre reportes de hogares que pertenecen a la misma vivienda: Solo al interior de 12 viv se presentan diferencias. <b>Ajuste efectuado:</b> Se decidió mantener los reportes por hogar en ENAHO.</p>
<p>u.o. – vivienda</p> <p>Tiene alumbrado eléctrico (si, no)</p>	<p>u.o. - hogar</p> <p>Cuál es el tipo de alumbrado que tiene su hogar - aceptar una o + alternativas – (electricidad, kerosene (mechero / lamparín), petróleo – gas (lámpara), vela, generador, otro, no utiliza)</p>	<p><b>Problema:</b> Censo reporta a nivel de vivienda (datos del hogar principal imputados a los secundarios) y la ENAHO a nivel de hogar. <i>Atenuante:</i> No se presentan diferencias importantes entre reportes de hogares que pertenecen a la misma vivienda: Solo al interior de 1 viv se presentaron diferencias. <b>Ajuste efectuado:</b> Se decidió mantener los reportes por hogar en ENAHO.</p>
<p>u.o. - vivienda</p> <p>Sin contar baño, cocina, ni pasadizo, ¿cuántas habitaciones en total tiene la vivienda?</p>	<p>u.o. - vivienda</p> <p>Sin contar baño, cocina, pasadizos ni garaje, ¿cuántas habitaciones en total tiene la vivienda?</p>	<p><b>Problema:</b> Los hogares secundarios en la ENAHO presentan missings (se encuesta solo al hogar principal). <b>Ajuste efectuado:</b> Se imputó el reporte del hogar principal a los hogares secundarios de ENAHO.</p>
<p>u.o. - vivienda</p> <p>Si hogar es el conjunto de personas que cocinan separadamente sus alimentos y duermen en la misma vivienda : ¿cuántos hogares ocupan la vivienda?</p>	<p>u.o. – vivienda</p> <p>Total de hogares que ocupan la vivienda</p>	<p><i>El tema de compatibilidad es tratado a profundidad en la primera parte del Anexo.</i></p>

**COMPARACIÓN DE VARIABLES DE LA ENAHO Y EL CENSO**

CENSO 93	ENAHO 98-4	AJUSTES EFECTUADOS
<p>u.o. – hogar</p> <p>Utiliza algún espacio de la vivienda para realizar una actividad que proporciones ingresos al hogar (sí, no)</p>	<p>u.o. – hogar</p> <p>Utiliza algún espacio de la vivienda para realizar una actividad que proporciones ingresos al hogar (sí, no)</p>	<p>Perfecto</p>
<p>u.o. – hogar</p> <p>Equipamiento del hogar</p> <p>Tiene el hogar alguno de los artefactos electrodomésticos siguientes:</p> <p style="padding-left: 40px;">Radio, tv blanco y negro, tv a color, equipo de sonido, video grabadora, lavadora de ropa, refrigeradora, aspiradora, lustradora</p> <p>Tiene el hogar alguna de las máquinas y equipos siguientes:</p> <p style="padding-left: 40px;">Máquina de coser, máquina de tejer, computadora, teléfono</p> <p>Tiene el hogar alguno de los medios de locomoción siguientes:</p> <p style="padding-left: 40px;">Automóvil o camioneta para uso particular, automóvil para trabajo, camioneta para trabajo, motocicleta, bicicleta, triciclo para trabajo</p>	<p>u.o. – hogar</p> <p>Su hogar tiene – aceptar una o + alt – (teléfono (fijo), celular, beeper)</p> <p>Equipamiento del hogar - Su hogar tiene:</p> <p style="padding-left: 40px;">Radio, tv blanco y negro, tv a color, refri – congeladora, máquina de coser, equipo de sonido, video grabadora, , lavadora, bicicleta, auto – camioneta, triciclo, motocicleta, computadora</p> <p style="padding-left: 40px;">Plancha, camión, mototaxi, otro</p>	<p><b>Problema:</b></p> <ol style="list-style-type: none"> <li>Diferencias entre categorías de equipamiento del hogar.</li> <li>No aparecen en ENAHO: aspiradora, lustradora, máquina de tejer.</li> </ol> <p><b>Ajuste efectuado:</b></p> <ol style="list-style-type: none"> <li>Coincidencia de categorías – equipamiento del hogar:</li> </ol> <p>Coinciden perfectamente:</p> <p style="padding-left: 40px;">radio, tv blanco y negro, tv a color, refri – congeladora, máquina de coser, equipo de sonido, video grabadora, , lavadora, bicicleta, motocicleta, computadora</p> <p>No coinciden perfectamente:</p> <ul style="list-style-type: none"> <li>▪ ENAHO: auto – camioneta, triciclo</li> </ul> <p>CENSO: automóvil o camioneta para uso particular, automóvil para trabajo, camioneta para trabajo, triciclo para trabajo</p> <p>=&gt; las categorías del censo fueron agrupadas en: “auto – camionet a” y “triciclo”</p> <ul style="list-style-type: none"> <li>▪ Las categ de acceso a servicio telefónico</li> </ul> <p>=&gt; las categorías de ENAHO fueron agrupadas en: “sí” (en caso de reportar “sí” en una o + de las categorías de tenencia de teléfono (teléfono fijo, celular), “no” (en caso de reportar “no” en todas las cateogías de tenencia de teléfono)</p> <ol style="list-style-type: none"> <li>Las variables que no estaban presentes en ambas cédulas, simultáneamente, fueron excluidas (aspiradora,</li> </ol>

**COMPARACIÓN DE VARIABLES DE LA ENAHO Y EL CENSO**

CENSO 93	ENAHO 98-4	AJUSTES EFECTUADOS
		lustradora, máquina de tejer, plancha, camión, mototaxi).
u.o. – individuo  Parentesco o relación con el (la) jefe (a) de hogar (jefe (a), esposa(o) – compañera (o), hijo(a), yerno – nuera, nieto(a), padres – suegros, otro pariente, trabajador del hogar, pensionista, otro no pariente)	u.o. – individuo  Relación de parentesco con el jefe de hogar (jefe (a), esposa(o) – compañera (o), hijo(a), yerno – nuera, padres – suegros, otro pariente, trabajador del hogar, pensionista, otro no pariente)	<b>Problema:</b> En ENAHO no aparece la categoría “nieto”. <b>Ajuste efectuado:</b> En el censo, se incluyó la categoría “nieto” en la de “otro pariente”.
u.o. – individuo  Sexo	u.o. – individuo  Sexo	Perfecto
u.o. – individuo  Edad (años cumplidos, para menores de 1 año - meses cumplidos)	u.o. – individuo  Edad (años cumplidos, para menores de 1 año - meses cumplidos)	Perfecto
u.o. – individuo de 5 años o +  Sabe leer y escribir (si, no)  u.o. – individuo de 5 años o +  Asistió o asiste a un centro educativo de enseñanza regular (no asiste pero asistió, asiste actualmente, nunca asistió)	u.o. – individuo de 3 años o +  Sabe leer y escribir (si, no)  u.o. – individuo de 3 años o +  Actualmente asiste a algún centro de enseñanza regular (si, no)	Perfecto  <i>En la encuesta, ningún niño menor de 5 años reportó saber leer ni escribir, por lo que no es necesario hacer correcciones por cota de edad.</i> <b>Problema:</b> 1. Solo la categoría “asiste actualmente” es idéntica para ambas bases. 2. Rangos de edad distintos. <b>8 Ajuste efectuado::</b> 1. Las categorías del Censo fueron agrupadas de la siguiente manera: Sí asiste actualmente a algún centro de enseñanza regular (“asiste actualmente”); No asiste actualmente a algún centro de enseñanza regular (“no asiste pero asistió” y “nunca asistió”).

**COMPARACIÓN DE VARIABLES DE LA ENAHO Y EL CENSO**

CENSO 93	ENAHO 98-4	AJUSTES EFECTUADOS
		2. Se acotó el rango de edad en ENAHO (5 años o +).
<p>u.o. – individuo de 5 años o +</p> <p>Cuál es el último año o grado de estudios o nivel que aprobó  <b>(ningún nivel, inicial o pre-escolar, primaria, secundaria, básica regular (EBR), básica laboral (EBL), superior no univ incompleta, superior no univ completa, superior univ incompleta, superior univ completa)</b>                      (incluye el número de años o grados aprobados de primaria, secundaria, EBR y EBL)</p>	<p>u.o. – individuo de 3 años o +</p> <p>Cuál es el último año o grado de estudios que aprobó                      (sin nivel, inicial, primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, superior no univ incompleta, superior no univ completa, superior univ incompleta, superior univ completa)                      (incluye número de años primaria, nivel y categoría público / privado)</p>	<p><b>Problema:</b></p> <ol style="list-style-type: none"> <li>Las categorías de primaria y secundaria de ENAHO no coinciden con las de primaria, secundaria, EBR y EBL del censo.</li> <li>Rango de edad distinto.</li> </ol> <p><b>Ajuste efectuado: (se mantiene el problema):</b>  <b>1.Las categorías fueron agrupadas de la siguiente manera:</b></p> <p>Primaria incompleta (primaria con 5 o – grados aprobados, EBR con 5 o – grados aprobados, EBL con 4 o – grados aprobados)</p> <p>Primaria completa (primaria con 6 o + grados aprobados, EBR con 6 grados aprobados, EBL con 5 grados aprobados)</p> <p>Secundaria incompleta (secundaria con 4 o – grados aprobados, EBR con 7 u 8 grados aprobados, EBL con 6, 7 u 8 grados aprobados)</p> <p>Secundaria completa (secundaria con 5 o + grados aprobados, EBR con 9 grados aprobados, EBL con 9 grados aprobados)</p> <p>2. Se acotó el rango de edad en ENAHO (5 años y +).</p>
<p>u.o. – individuo de 6 años o +</p> <p>La semana pasada se encontraba (trabajando por algún ingreso, no trabajó pero tenía trabajo, ayudando a un familiar sin pago alguno, buscando trabajo – habiendo trabajado antes, buscando trabajo por primera vez, estuvo al cuidado de su hogar y no trabajó, estudiando y no trabajó, vivía de su pensión o jubilación y no trabajó, otra)</p>	<p>u.o. – individuo de 14 años o +</p> <p>La semana pasada (trabajó, buscó trabajo, no trabajó ni buscó trabajo)</p>	<p><b>Problema:</b></p> <ol style="list-style-type: none"> <li>Las categorías no coinciden y la cota de edad tampoco.</li> </ol> <p><b>Ajuste efectuado:</b></p> <ol style="list-style-type: none"> <li>Las categorías fueron agrupadas de la siguiente manera:</li> </ol> <p>Trabajó (trabajó por algún ingreso, no trabajó pero tenía trabajo, ayudando a un familiar sin pago alguno)</p> <p>Buscó trabajo (buscando trabajo – habiendo</p>

**COMPARACIÓN DE VARIABLES DE LA ENAHO Y EL CENSO**

CENSO 93	ENAHO 98-4	AJUSTES EFECTUADOS
		trabajando antes, buscando trabajo por primera vez) No buscó ni trabajó (estuvo al cuidado de su hogar y no trabajó, estudiando y no trabajó, vivía de su pensión o jubilación y no trabajó, otra (las personas que entran en esta categ no son entrevistadas en relación a ocupaciones anteriores o actuales)) 2. Se acotó el rango de edad en el Censo (14 años o +).
u.o. – individuo de 6 años o +  Cuál es la ocupación principal que desempeñó	u.o. – individuo de 14 años o +  Cuál es la ocupación principal que desempeñó	<b>Problema:</b> Rangos de edad distintos. <b>Ajuste efectuado:</b> Se acotó el rango de edad en el Censo (14 años o +).
u.o. – individuo de 6 años o +  A qué se dedica el negocio, organismo o empresa en la que trabajó	u.o. – individuo de 14 años o +  A qué se dedica el negocio, organismo o empresa en la que trabajó	<b>Problema:</b> Rangos de edad distintos. <b>Ajuste efectuado:</b> Se acotó el rango de edad en el Censo (14 años o +).
u.o. – individuo de 6 años o +  En su centro de trabajo usted era (obrero, empleado, trabajador familiar no remunerado, trabajador del hogar, trabajador independiente o por cuenta propia, empleador o patrono)	u.o. – individuo de 14 años o +  En su centro de trabajo usted era (empleador o patrono, trabajador independiente, empleado, obrero, trabajador familiar no remunerado, trabajador del hogar, otro)	<b>Problema:</b> Rangos de edad distintos. <b>Ajuste efectuado:</b> Se acotó el rango de edad en el Censo (14 años o +).
u.o. – individuo de 12 años o +  Su estado civil es (conviviente, casado, viudo, divorciado, separado, soltero)	u.o. – individuo de 12 años o +  Su estado civil o conyugal es (conviviente, casado, viudo, divorciado, separado, soltero)	Perfecto

### ANEXO 3

#### Variables consideradas en la Formulación de los Modelos

Código	Nombre	Definición de la variable
v1	miemhog	número de miembros del hogar
v2	tothog	número de hogares que residen en la vivienda
v3	electric	el hogar tiene alumbrado eléctrico
v4	habingr	el hogar utiliza algún espacio de la vivienda para actividad que genere ingresos
v5	radio	el hogar tiene radio
v6	tvblnegr	el hogar tiene tv blanco y negro
v7	tvcolor	el hogar tiene tv a color
v8	refrig	el hogar tiene refrigeradora / congeladora
v9	maqcoser	el hogar tiene máquina de coser
v10	eqsonido	el hogar tiene equipo de sonido
v11	videogr	el hogar tiene video grabadora
v12	lavadora	el hogar tiene lavadora
v13	biciclet	el hogar tiene bicicleta
v14	autocam	el hogar tiene auto o camioneta
v15	triciclo	el hogar tiene triciclo
v16	motocicl	el hogar tiene motocicleta
v17	comput	el hogar tiene computadora
v18	ntrabhog	número de trabajadores del hogar
v19	npension	número de pensionistas del hogar
v20	sexjef	sexo del jefe de hogar
v21	pmiemuj	proporción de mujeres respecto del total de miembros del hogar, incluyendo al jefe de hogar
v22	edadjef	edad del jefe de hogar
v23	pmuj0a13	proporción de mujeres de 0 a 13 años respecto del total de miembros, incluyendo al jefe de hogar
v24	phom0a13	proporción de hombres de 0 a 13 años respecto del total de miembros, incluyendo al jefe de hogar
V25	pmuj65	proporción de mujeres mayores de 65 años respecto del total de miembros, incluye al jefe de hogar
v26	phom65	proporción de hombres mayores de 65 años respecto del total de miembros, incluye al jefe de hogar
v27	panalfab	proporción de analfabetos de 14 años o +, incluyendo al jefe de hogar
v28	pasisten	proporción de miembros de 5 años o + que asisten actualmente a algún centro de enseñanza (respecto del total de miembros de 5 años y +), incluyendo al jefe de hogar
v29	edusinj	jefe del hogar sin nivel educativo (sin nivel o con ed inicial)
v30	edjprinc	jefe del hogar con ed primaria incompleta
v31	edjprcom	jefe del hogar con ed primaria completa
v32	edjseinc	jefe del hogar con ed secundaria incompleta
v33	edjsecom	jefe del hogar con ed secundaria completa
v34	edjsnuin	jefe del hogar con ed superior no universitaria incompleta
v35	edjsnuco	jefe del hogar con ed superior no universitaria completa
v36	edjsuinc	jefe del hogar con ed superior universitaria incompleta
v37	pmedusin	proporción de miembros del hogar de 5 años y + sin nivel educativo (sin nivel o con ed inicial), excluyendo al jefe de hogar, respecto del total de miembros del hogar
v38	pmedupri	proporción de miembros del hogar de 5 años y + con ed primaria incompleta, excluyendo al jefe de hogar, respecto del total de miembros del hogar
v39	pmedprco	proporción de miembros del hogar de 5 años y + con ed primaria completa, excluyendo al jefe de hogar, respecto del total de miembros del hogar"

Código	Nombre	Definición de la variable
v40	pmedsein	proporción de miembros del hogar de 5 años y + con ed secundaria incompleta, excluyendo al jefe de hogar, respecto del total de miembros del hogar
v41	pmedseco	proporción de miembros del hogar de 5 años y + con ed secundaria completa, excluyendo al jefe de hogar, respecto del total de miembros del hogar
v42	pmesnuin	proporción de miembros del hogar de 5 años y + con ed superior no universitaria incompleta, excluyendo al jefe de hogar, respecto del total de miembros del hogar
v43	pmesnuco	proporción de miembros del hogar de 5 años y + con ed superior no universitaria completa, excluyendo al jefe de hogar, respecto del total de miembros del hogar
v44	pmesuinc	proporción de miembros del hogar de 5 años y + con ed superior universitaria incompleta, excluyendo al jefe de hogar, respecto del total de miembros del hogar
v45	pmesucom	proporción de miembros del hogar de 5 años y + con ed superior universitaria completa, excluyendo al jefe de hogar, respecto del total de miembros del hogar
v46	pmieocum	proporción de mujeres ocupadas de 14 años y + respecto del total de miembros del hogar, incluyendo al jefe de hogar
v47	pagrotot	proporción de miembros de 14 años y +, incluyendo al jefe de hogar, que trabajan en un negocio agrícola, pecuario, de silvicultura o caza, respecto del total de miembros ocupados del hogar
v48	pcomtot	proporción de miembros de 14 años y +, incluyendo al jefe de hogar, que trabajan en un negocio comercial, respecto del total de miembros ocupados del hogar
v49	pcota1	proporción de miembros, incluyendo al jefe de hogar, cuya ocupación es la de ejecutivos, profesionales, técnicos, oficinistas, servicios, otros (FFAA y policiales)
v50	vvalquil	la vivienda es alquilada
v51	vvpropia	la vivienda es propia - totalmente pagada o comprándola a plazos
v52	numhab	número de habitaciones de la vivienda, son contar con baño, cocina, pasadizos
v53	prladril	las paredes de la vivienda son de ladrillo o bloque de cemento
v54	prpiedra	las paredes de la vivienda son de piedra o sillar con cal o cemento
v55	pradobe	las paredes de la vivienda son de adobe o tapia
v56	prquinch	las paredes de la vivienda son de quincha
v57	prbarro	las paredes de la vivienda son de piedra con barro
v58	prmadera	las paredes de la vivienda son de madera
v59	prestera	las paredes de la vivienda son de estera
v60	parquet	el material predominante en los pisos es parquet o madera pulida
v61	lamina	el material predominante en los pisos es lámina asfáltica, vinílico o similares
v62	losetas	el material predominante en los pisos es loseta, terrazos o similares
v63	madera	el material predominante en los pisos es madera
v64	cemento	el material predominante en los pisos es cemento
v65	tierra	el material predominante en los pisos es tierra
v66	aguaenvv	el abastecimiento de agua procede de red pública dentro de la vivienda
v67	higenvv	el servicio higiénico está conectado a red pública dentro de la vivienda
v68	telef	el hogar tiene teléfono

**Listado de Variables de los Modelos Estimados**  
(nemónicos)

Modelo Lineal Con Interacciones		Modelo Probit Con Interacciones	Modelo Lineal Sin Interacciones	Modelo Probit Sin Interacciones
<i>Costa Urbana</i>				
v1	v20v24	v1	v1	v1
v11	v20v34	v11	v11	v11
v16	v20v38	v15	v16	v15
v17	v20v39	v18	v17	v18
v20	v21v33	v23	v20	v23
v21	v21v59	v24	v21	v24
v23	v23v24	v27	v23	v27
v24	v23v29	v29	v24	v29
v28	v23v36	v3	v28	v3
v29	v24v30	v38	v29	v38
v3	v29v38	v42	v3	v42
v30	v30v60	v53	v30	v53
v31	v31v40	v55	v31	v55
v32	v31v45	v58	v32	v58
v33	v31v59	v59	v33	v59
v34	v31v68	v6	v34	v6
v35	v32v45	v65	v35	v65
v36	v32v53	v68	v36	v68
v37	v32v59	v7	v37	v7
v38	v34v37	v9	v38	v9
v39	v34v39	v15v29	v39	
v40	v34v45	v15v55	v40	
v45	v35v39	v1v1	v45	
v52	v35v60	v1v27	v52	
v53	v36v37	v1v53	v53	
v56	v36v59	v1v55	v56	
v57	v39v45	v23v59	v57	
v59	v3v16	v24v24	v59	
v6	v3v30	v27v29	v6	
v60	v3v37	v27v42	v60	
v68	v3v38	v27v53	v68	
v7	v3v68	v27v59	v7	
v8	v3v7	v29v38	v8	
v9	v40v52	v3v9	v9	
v11v68	v45v56	v42v3		
v16v34	v59v68	v42v55		
v17v20	v6v33	v59v9		
v17v24	v6v34	v65v3		
v17v31	v6v40	v68v38		
v17v32	v6v56	v68v42		
v17v33	v6v60	v68v7		
v17v35	v8v28	v6v55		
v17v39	v8v29			
v17v53	v8v33			
v17v60	v8v37			
v17v68	v9v30			

<b>Modelo Lineal Con Interacciones</b>	<b>Modelo Probit Con Interacciones</b>	<b>Modelo Lineal Sin Interacciones</b>	<b>Modelo Probit Sin Interacciones</b>
v1v1 v1v34 v1v59	v9v40 v9v59		
<b>Costa Rural</b>			
v1 v11 v23 v24 v27 v29 v30 v31 v39 v44 v5 v50 v52 v53 v59 v6 v60 v64 v67 v68 v7 v11v50 v11v64 v1v1 v1v53 v1v67 v23v24 v24v68	v27v6 v27v67 v29v53 v29v59 v30v44 v30v67 v31v39 v39v64 v39v67 v44v5 v44v52 v44v67 v50v53 v50v6 v50v64 v52v59 v52v67 v53v68 v59v6 v5v50 v5v68 v6v64 v6v68	v1 v12 v23 v24 v27 v30 v39 v45 v47 v5 v50 v52 v53 v57 v58 v59 v6 v64 v7 v1v23 v1v24 v1v30 v1v47 v1v7 v23v24 v23v27 v47v52 v6v27	v1 v11 v23 v24 v27 v29 v30 v31 v39 v47 v44 v5 v50 v52 v53 v57 v58 v59 v6 v60 v64 v67 v68 v7
<b>Sierra Urbana</b>			
v1 v10 v11 v14 v16 v17 v23 v24 v25 v27 v28 v3 v30 v31	v23v40 v23v44 v23v47 v23v61 v24v25 v24v28 v24v3 v24v30 v24v34 v24v44 v24v52 v24v6 v24v61 v25v31	v1 v10 v12 v13 v14 v18 v20 v23 v24 v27 v29 v3 v30 v39	v1 v10 v11 v14 v16 v17 v23 v24 v25 v27 v28 v3 v30 v31

Modelo Lineal Con Interacciones		Modelo Probit Con Interacciones	Modelo Lineal Sin Interacciones	Modelo Probit Sin Interacciones
v34	v25v38	v46	v34	v46
v38	v25v39	v47	v38	v47
v39	v25v41	v52	v39	v52
v40	v25v47	v63	v40	v63
v41	v25v6	v68	v41	v68
v42	v25v61	v7	v42	v7
v43	v25v68	v8	v43	v8
v44	v27v3	v1v1	v44	
v47	v27v31	v1v23	v47	
v49	v27v39	v24v46	v49	
v52	v27v49	v29v13	v52	
v6	v28v44	v30v47	v6	
v61	v28v47	v3v63	v61	
v65	v28v49	v52v10	v65	
v67	v28v65	v52v39	v67	
v68	v30v39	v68v30	v68	
v7	v30v41	v68v52	v7	
v10v11	v30v44	v7v46		
v10v14	v31v67	v7v52		
v10v23	v31v7	v8v39		
v10v25	v34v38			
v10v28	v34v40			
v10v52	v34v42			
v11v14	v34v43			
v11v3	v34v44			
v11v34	v34v47			
v11v41	v34v49			
v11v61	v34v7			
v14v3	v38v6			
v14v34	v39v7			
v14v41	v3v34			
v14v52	v3v43			
v14v65	v3v44			
v16v3	v40v40			
v16v30	v40v42			
v16v31	v40v6			
v16v39	v40v61			
v16v40	v40v68			
v16v44	v41v68			
v16v65	v42v7			
v17v42	v43v47			
v17v49	v43v49			
v17v52	v44v52			
v1v1	v49v49			
v1v23	v49v6			
v1v25	v52v52			
v1v28	v52v7			
v1v34	v6v67			
v23v25				

Modelo Lineal Con Interacciones		Modelo Probit Con Interacciones	Modelo Lineal Sin Interacciones	Modelo Probit Sin Interacciones
<i>Sierra Rural</i>				
v1	v24v53	v1	v1	v1
v10	v24v56	v13	v10	v13
v12	v27v35	v14	v12	v14
v13	v27v67	v15	v13	v15
v18	v28v28	v17	v18	v17
v2	v28v36	v23	v2	v23
v23	v32v38	v24	v23	v24
v24	v32v40	v27	v24	v27
v27	v32v56	v28	v27	v28
v28	v32v65	v33	v28	v33
v32	v33v53	v35	v32	v35
v33	v35v39	v4	v33	v4
v35	v35v67	v44	v35	v44
v36	v36v37	v45	v36	v45
v37	v36v49	v47	v37	v47
v38	v36v65	v49	v38	v49
v39	v37v65	v5	v39	v5
v4	v38v39	v52	v4	v52
v40	v39v53	v6	v40	v6
v49	v40v53	v7	v49	v7
v5	v40v56	v8	v5	v8
v52	v49v56	v9	v52	v9
v53	v4v23	v1v1	v53	
v56	v4v24	v1v15	v56	
v6	v4v67	v23v13	v6	
v61	v53v65	v23v24	v61	
v65	v5v35	v23v28	v65	
v66	v5v53	v23v9	v66	
v67	v5v67	v27v13	v67	
v7	v5v7	v27v15	v7	
v8	v65v66	v27v28	v8	
v9	v6v67	v27v47	v9	
v10v35	v6v7	v4v24		
v10v40	v7v18	v5v4		
v10v53	v7v23	v6v5		
v12v52	v7v32	v7v13		
v12v65	v7v36	v9v15		
v13v24	v7v38			
v13v66	v7v56			
v18v67	v7v65			
v1v1	v7v67			
v1v27	v8v28			
v1v4	v8v33			
v1v67	v8v35			
v23v27	v8v39			
v23v28	v8v49			
v23v56	v9v23			
v24v39	v9v35			
v24v40	v9v56			

Modelo Lineal Con Interacciones		Modelo Probit Con Interacciones	Modelo Lineal Sin Interacciones	Modelo Probit Sin Interacciones
<i>Selva Urbana</i>				
v1	v12v21	v1	v1	v1
v10	v12v24	v14	v10	v14
v11	v12v43	v21	v11	v21
v12	v12v47	v24	v12	v24
v15	v12v51	v27	v15	v27
v17	v15v41	v29	v17	v29
v20	v15v47	v34	v20	v34
v21	v15v51	v36	v21	v36
v23	v15v67	v37	v23	v37
v24	v17v43	v41	v24	v41
v29	v17v49	v42	v29	v42
v3	v1v1	v46	v3	v46
v30	v1v17	v47	v30	v47
v31	v20v53	v49	v31	v49
v41	v21v24	v53	v41	v53
v43	v21v61	v6	v43	V6
v45	v23v63	v61	v45	v61
v47	v24v63	v65	v47	v65
v48	v29v63	v67	v48	v67
v49	v30v43	v68	v49	v68
v51	v30v45	v1v21	v51	
v53	v3v15	v24v24	v53	
v54	v3v20	v27v27	v54	
v57	v3v21	v41v46	v57	
v61	v3v41	v47v47	v61	
v63	v41v49	v49v21	v63	
v67	v43v48	v49v46	v67	
v68	v43v49	v65v21	v68	
v7	v43v61	v65v53	v7	
v10v15	v47v68			
v10v48	v49v49			
v10v61	v49v51			
v11v12	v51v54			
v11v23	v53v63			
v11v24	v54v63			
v11v30	v67v68			
v11v31	v7v11			
v11v45	v7v15			
v11v48	v7v21			
v11v49	v7v24			
v12v17	v7v48			
<i>Selva Rural</i>				
v1		v1	v1	v1
v10		v13	v10	v13
v13		v16	v13	v16
v14		v23	v14	v23
v15		v24	v15	v24
v16		v28	v16	v28

<b>Modelo Lineal Con Interacciones</b>	<b>Modelo Probit Con Interacciones</b>	<b>Modelo Lineal Sin Interacciones</b>	<b>Modelo Probit Sin Interacciones</b>
v23	v29	v23	v29
v24	v3	v24	v3
v27	v30	v27	v30
v28	v31	v28	v31
v3	v32	v3	v32
v33	v34	v33	v34
v35	v35	v35	v35
v36	v37	v36	v37
v39	v39	v39	v39
v45	v4	v45	v4
v51	v47	v51	v47
v53	v48	v53	v48
v54	v51	v54	v51
v62	v53	v62	v53
v66	v66	v66	v66
v10v33	v9		v9
v10v35	v13v32		
v10v66	v16v30		
v13v15	v24v37		
v14v35	v48v66		
v14v51	v4v28		
v14v53	v4v51		
v15v23	v51v47		
v15v27			
v15v53			
v16v33			
v16v53			
v1v1			
v1v16			
v23v53			
v28v35			
v35v39			
v35v66			
v39v51			
v51v54			
v53v66			
<b>Lima Metropolitana</b>			
v1	v14v47	v1	v1
v10	v17v31	v10	v10
v11	v17v32	v14	v14
v12	v17v39	v16	v16
v14	v17v65	v32	v32
v17	v1v1	v36	v36
v2	v1v14	v42	v42
v23	v1v33	v44	v44
v24	v1v41	v46	v46
v25	v1v47	v5	v5
v28	v1v59	v53	v53
v29	v23v41	v62	v62

<b>Modelo Lineal Con Interacciones</b>		<b>Modelo Probit Con Interacciones</b>	<b>Modelo Lineal Sin Interacciones</b>	<b>Modelo Probit Sin Interacciones</b>
v30	v24v25	v66	v30	v66
v31	v24v32	v68	v31	v68
v32	v24v38	v8	v32	v8
v33	v24v46	v1v1	v33	
v35	v25v33	v1v32	v35	
v38	v28v47	v68v46	v38	
v39	v28v59		v39	
v40	v29v64		v40	
v41	v2v12		v41	
v46	v2v30		v46	
v47	v2v31		v47	
v53	v2v38		v53	
v54	v2v40		v54	
v57	v30v41		v57	
v59	v31v47		v59	
v64	v31v53		v64	
v65	v32v41		v65	
v68	v32v46		v68	
v8	v32v64		v8	
v10v17	v33v40			
v10v47	v35v39			
v10v64	v41v65			
v11v39	v53v68			
v11v53	v8v11			
v11v68	v8v12			
v12v17	v8v17			
v12v25	v8v23			
v12v32	v8v29			
v12v40	v8v47			
v14v30	v8v59			