
POBREZA Y DESIGUALDAD EN AMERICA LATINA:
CONCEPTOS, HERRAMIENTAS Y APLICACIONES

Capítulo 3

**VARIABLES Y FUENTES DE
INFORMACIÓN**

Esta versión: 27 de septiembre, 2010 *

* Este documento es un borrador del capítulo 3 del libro “Pobreza y Desigualdad en América Latina. Conceptos, herramientas y aplicaciones” por Leonardo Gasparini, Martín Cicowiez y Walter Sosa Escudero. El libro se realiza en el marco del CEDLAS, el Centro de Estudios Distributivos, Laborales y Sociales de la Universidad Nacional de La Plata (cedlas.econo.unlp.edu.ar). Por favor, no citar sin permiso. Se agradecen los comentarios.

Índice del Capítulo 3

1. INTRODUCCIÓN	3
2. VARIABLES DE INTERÉS	4
3. PERSONAS Y HOGARES	18
4. EDAD Y COMPARACIONES INTERTEMPORALES	30
5. PERSONAS, FACTORES Y GRUPOS	36
6. FUENTES DE INFORMACIÓN.....	39
7. INGRESO Y CONSUMO EN LAS ENCUESTAS	56
APÉNDICE: EN LA PRÁCTICA.....	64

Pobreza y desigualdad en América Latina : conceptos, herramientas y aplicaciones / Leonardo Gasparini; Martín Cicowiez; Walter Sosa Escudero. - 1a ed. - La Plata : Universidad Nacional de La Plata, 2010.

CD-ROM.

ISBN 978-950-34-0667-0

1. Problemas Sociales. 2. Pobreza. 3. Enseñanza Universitaria. I. Cicowiez, Martín II. Sosa Escudero, Walter III. Título

CDD 362.5

Fecha de catalogación: 20/08/2010

1. Introducción

En el capítulo anterior estudiamos la distribución de una variable de bienestar individual a la cual no definimos con precisión, ni discutimos conceptualmente. Este capítulo está destinado a cubrir estos aspectos, para lo cual presentamos las ventajas y limitaciones de un conjunto alternativo de variables que potencialmente pueden ocupar el centro del análisis distributivo. Una vez escogida alguna variable relevante, ésta debe ser estimada a partir de datos concretos de la realidad. El segundo objetivo de este capítulo es listar y caracterizar las fuentes de información que proveen los datos para un estudio distributivo típico, en especial aquellas disponibles en América Latina.

El capítulo comienza con un debate conceptual fundamental que ha merecido la atención de filósofos e investigadores sociales: ¿cuál es la variable (o las variables) de la cual nos interesa estudiar su distribución?, ¿qué razones positivas y normativas hay detrás de este interés? En la sección 2 damos una primera mirada a esta discusión, que extenderemos en los siguientes capítulos del libro. Pese a que el debate sobre el objeto de análisis distributivo es muy rico, se suele dirimir por razones prácticas: la disponibilidad de información y la sencillez del análisis empírico. Por claras ventajas en estas dos dimensiones, el ingreso y el consumo monetario aparecen como las dos variables sobre las que se concentra la literatura distributiva empírica.

La sección 3 reconoce que las personas no viven solas, sino que pertenecen a hogares en los que se comparten recursos, lo cual genera un conjunto de nuevas preguntas. ¿Debemos comparar ingresos entre personas o entre hogares? Dado que las necesidades pueden diferir entre personas por factores como la edad o el género, ¿nos interesa la distribución del ingreso (o del consumo) total, del ingreso per cápita o del ingreso ajustado por alguna escala de equivalencia?

Los ingresos y consumos no son uniformes a lo largo de la vida, sino que por lo general suelen crecer hasta la edad adulta y luego caer, dependiendo de la cobertura del sistema de seguridad social. Conforme a este patrón existen argumentos que sostienen la necesidad de comparar los niveles de vida de personas sólo de la misma edad, o alternativamente comparar los ingresos intertemporales o permanentes de las personas, es decir aquellos obtenidos a lo largo de toda la vida, y no los de un momento particular. Estos argumentos son abordados en la sección 4 del capítulo.

Si bien gran parte de la literatura distributiva, incluido este libro, estudia la distribución del bienestar entre personas, alternativamente es posible dirigir el análisis hacia la distribución entre otras unidades. La sección 5 examina brevemente la distribución del ingreso entre factores productivos (distribución *funcional*) y entre grupos definidos por alguna característica (ej. área geográfica).

Este es un libro eminentemente empírico, por lo que la discusión sobre las fuentes de información y la calidad de los datos ocupa un lugar importante. La sección 6 está asignada a este punto; en particular, se discuten las características de las encuestas de hogares, sus ventajas y limitaciones, con especial énfasis en las encuestas disponibles en América Latina.

Finalmente, en la sección 7 se explican los pasos a seguir en la generación de variables de ingreso y consumo a partir de las encuestas de hogares de la región, y se comentan algunos de los principales problemas en el trabajo empírico. Como en todo el libro este capítulo incluye un apéndice, donde se guía al lector hacia la implementación práctica de lo discutido en las distintas secciones.

2. Variables de interés

El término *distribución del ingreso* es utilizado para englobar un conjunto de fenómenos relacionados con la equidad económica. Así, el principal compendio de Economía sobre este tema es llamado *Handbook of Income Distribution*, aunque desde su capítulo 1 se alerta sobre la “necesidad de liberar el análisis de la desigualdad económica al confinamiento del espacio del ingreso o la posesión de bienes” (Sen, 2000). Si bien en la práctica el ingreso ocupa un lugar central en el análisis distributivo, es importante detenernos en pensar el problema conceptualmente y evaluar alternativas.

Si nos motiva el interés por la equidad, ¿cuál es la variable cuya distribución deberíamos conocer y estudiar? ¿Nos interesa la pobreza como carencia de qué variable o conjunto de variables? ¿Nos preocupa la desigualdad entre las personas en términos de cuáles factores? Estas son preguntas conceptuales intensamente debatidas en las ciencias sociales y en la filosofía, sobre las que existe una vasta literatura.¹ Amartya Sen, economista y filósofo, premio Nobel de Economía en 1998, sostiene que todas las concepciones de equidad implican la igualdad de algo, pero difieren en la variable que consideran relevante igualar. A riesgo de simplificar demasiado una discusión muy rica, es posible identificar dos concepciones básicas de equidad: una basada en resultados y otra en oportunidades.

2.1. Resultados y oportunidades

Los resultados son consecuencia, al menos parcialmente, de la acción deliberada de las personas. Las oportunidades, en cambio, son circunstancias dadas para el individuo, que limitan el conjunto de resultados a los cuales puede tener acceso a partir de sus elecciones. El ingreso individual, por ejemplo, es una variable de resultado, en cuanto al menos en parte es determinado por decisiones de la persona respecto de su esfuerzo, capacitación, toma de riesgos, y diversas elecciones laborales y familiares. En cambio, el ingreso de los padres es una variable que forma parte del conjunto de oportunidades de un niño, ya que afecta sus posibilidades económicas futuras pero está fuera de su control.

Las concepciones de equidad basadas en resultados y oportunidades son discutidas más extensamente en el capítulo 6. Adelantamos dos de las principales conclusiones. Primero, la idea de igualdad de oportunidades resulta conceptualmente más atractiva, y

¹ Para una introducción desde la Economía consultar Le Grand (1991), Roemer (1998) y Sen (1992).

de hecho es la más aceptada por la opinión pública. Existen diferencias de resultados que si provienen de decisiones tomadas libremente en un escenario de igualdad de oportunidades, no son evaluadas por la mayoría de las personas como injustas. Segundo, desafortunadamente la implementación empírica del concepto de oportunidades es engorrosa y ambigua. La idea de oportunidades no es fácil de definir, no tiene un correlato empírico claro y es multidimensional, a diferencia de las típicas variables de resultado. Por esta razón, la enorme mayoría de los estudios sobre equidad se concentran en la distribución del ingreso, el consumo u otra variable de resultado, en lugar de focalizarse en conceptos más ambiciosos, como el de oportunidades. Gran parte de este libro sigue este camino, dictado más por la conveniencia práctica que por el convencimiento teórico.²

2.2. Utilidad y aproximaciones monetarias

Supongamos que decidimos trabajar con la distribución de alguna variable de resultado y no de oportunidades. Un primer candidato es la utilidad (o bienestar, o felicidad) individual. Gran parte del análisis económico se centra, de hecho, en esa variable. Un supuesto crucial que caracteriza al paradigma económico tradicional es que las personas buscan maximizar su utilidad actuando racionalmente en un mundo con restricciones presupuestarias, de tiempo e informativas.

Tomemos el modelo más sencillo de un solo período sin incertidumbre, asumiendo que las personas viven solas. En este mundo cada individuo es un hogar, y el ingreso de la persona, al que llamamos y , es trivialmente el ingreso total y el ingreso per cápita de su hogar unipersonal. Más adelante introduciremos las complicaciones que surgen en la vida real, donde las personas se agrupan en familias. En este modelo simple cada persona busca resolver el siguiente problema:

$$(2.1) \quad \text{Max}_q u(q) \text{ sujeto a } p \cdot q = y$$

donde u es la función de utilidad individual que depende del consumo de un vector de bienes y servicios q , disponibles en el mercado a precios fijos p .³ De la maximización de la utilidad surgen funciones de demanda de bienes $q(p,y)$, las que reemplazadas en la función objetivo $u(q)$ generan la llamada función de utilidad indirecta $v(p,y)$. Estas funciones indican la máxima utilidad a la que puede acceder una persona que enfrenta precios p y dispone de un ingreso y . Una alternativa para el análisis distributivo sería comparar los valores de utilidad v entre las personas. De hecho, buena parte de los resultados de la Economía del Bienestar provienen de resolver la maximización de una

² La literatura que intenta hacer operativo el concepto de igualdad de oportunidades es revisada en el capítulo 7.

³ Nótese que el ingreso y aparece en este modelo sencillo como dado, y por ende, como determinante de las oportunidades para alcanzar niveles de utilidad. En el mundo real o en modelos más elaborados, y tal como discutimos antes, el ingreso no está dado, sino que depende de un conjunto de decisiones, por lo que se trata de una variable de resultado y no de oportunidades.

función de bienestar social W cuyos argumentos son precisamente las utilidades individuales. A esta función $W(u_1(q_1) \dots, u_N(q_N)) = W(v_1(p, y_1) \dots, v_N(p, y_N))$ se la conoce habitualmente como función de bienestar social de Bergson-Samuelson, y la usaremos extensamente en los capítulos siguientes. Se argumenta que la maximización de esta función debería ser la preocupación central de las políticas públicas.

Si bien la utilidad individual parece un candidato natural para centrar los estudios distributivos, esta variable enfrenta algunos problemas prácticos fundamentales: no es observable, y es difícil de estimar y de comparar entre personas. Una alternativa es concentrarse en el vector de consumo q , ignorando la función u que transforma cantidades de bienes en utilidad. Sin embargo, una vez más nos topamos con el problema de la multidimensionalidad. El vector q es una lista potencialmente extensa de bienes y servicios que generan utilidad. La comparación de vectores es engorrosa y ambigua: si la persona A consume 2 manzanas y 3 peras, y la persona B consume 3 peras y 2 manzanas, ¿quién goza de un nivel de vida más alto?

Frente a estos problemas la alternativa que ha seguido la Economía es concentrarse en alguna de las dos variables monetarias del problema de maximización planteado anteriormente: el ingreso y , o el gasto de consumo $c \equiv p \cdot q$, que en este ejemplo sencillo coinciden. Estas dos variables tienen la gran ventaja de ser escalares (es decir, números en lugar de vectores) expresados en unidades monetarias (pesos). Conceptualmente tienen también el atractivo de representar la base a partir de la cual cada persona deriva utilidad de acuerdo a su función particular. A menudo se sostiene que el gasto de consumo y el ingreso son aproximaciones al “nivel de vida” de la persona, entendido éste como el conjunto de bienes y servicios a los que la persona tiene acceso.

Existen razones teóricas que justifican la elección del ingreso o el consumo como aproximaciones monetarias al nivel de utilidad. Los párrafos que siguen revisan esos argumentos con el lenguaje técnico de la Economía, por lo que pueden ser saltados por el lector no interesado.⁴

El enfoque más popular, iniciado por Samuelson (1974), es el de la *utilidad monetaria* (*money metric utility*) y consiste en calcular el costo monetario de alcanzar un determinado nivel de utilidad, dado un conjunto de precios de referencia. La utilidad monetaria para la persona i (u_i^m) es

$$(2.2) \quad u_i^m = e(u_i, p_0)$$

donde $e(\cdot)$ es la función de gasto indirecta, por lo que u_i^m indica el costo monetario de alcanzar el nivel de utilidad u_i a los precios de referencia p_0 . Esta función es difícil de estimar ya que requiere conocer las preferencias u , pero puede ser aproximada por una expansión de Taylor de primer orden,

$$(2.3) \quad e(u_i, p_0) \approx e(u_i, p_i) + (p_0 - p_i) \cdot \frac{\partial e(u_i, p_i)}{\partial p}$$

⁴ La exposición sigue a Deaton y Muellbauer (1980) y Deaton y Zaidi (2003).

donde p_i es el vector de precios relevante para i dado su vector de consumo (y por lo tanto variable en el tiempo y en el espacio). El *lemma de Shephard* indica que en el óptimo la derivada de la función de gasto indirecta con respecto al vector de precios es el vector de cantidades efectivamente consumidas q_i . Adicionalmente, dado que se asume que i maximiza utilidad, entonces $e(u_i, p_i) = p_i \cdot q_i$. Combinando estas ecuaciones,

$$(2.4) \quad u_i^m = e(u_i, p_0) \approx p_0 \cdot q_i$$

Es posible rescribir a $p_0 \cdot q_i$ como

$$(2.5) \quad p_0 \cdot q_i = \frac{p_i \cdot q_i}{\left(\frac{p_i \cdot q_i}{p_0 \cdot q_i} \right)} = \frac{p_i \cdot q_i}{I_{p_i}}$$

donde I_{p_i} es un índice de precios tipo Paasche específico del individuo i . Combinando las ecuaciones (2.4) y (2.5),

$$(2.6) \quad u_i^m \approx \frac{p_i \cdot q_i}{I_{p_i}} = \frac{c_i}{I_{p_i}}$$

La utilidad monetaria de i puede ser aproximada por el gasto monetario total del individuo dividido por un índice Paasche que capte los precios relativos que enfrenta la persona respecto del vector de precios de referencia. Dado que en este modelo sencillo el ingreso es igual al consumo, y_i/I_{p_i} es también un indicador de bienestar válido.

El enfoque alternativo es el del cociente de bienestar (*welfare ratio*) y fue propuesto por Blackorby y Donaldson (1997).⁵ Este cociente surge de dividir el gasto del individuo por el gasto requerido para alcanzar un nivel de utilidad de referencia a los precios efectivos que enfrenta. En la práctica, como referencia se suele tomar un nivel de utilidad bajo, identificado con alguna idea de pobreza (u_z). El ratio de bienestar de i es entonces

$$(2.7) \quad r_i = \frac{e(u_i, p_i)}{e(u_z, p_i)}$$

Para llevar este coeficiente a valores monetarios lo usual es multiplicarlo por el gasto necesario para alcanzar un nivel básico de utilidad $e(u_z, p_0)$.⁶ El valor monetario de la utilidad bajo este método, denotado como u_i^r , es entonces

$$(2.8) \quad u_i^r = \frac{e(u_i, p_i)}{e(u_z, p_i)} \cdot e(u_z, p_0)$$

⁵ El enfoque alternativo surge ante el problema de que la utilidad monetaria no es una función cóncava del gasto o el ingreso, lo que puede generar que políticas redistributivas en el espacio del ingreso no necesariamente reduzcan la desigualdad en el espacio de la utilidad monetaria.

⁶ Este gasto está estrechamente emparentado con la línea de pobreza, como veremos en el capítulo siguiente.

o, de forma equivalente

$$(2.9) \quad u_i^r = \frac{P_i \cdot q_i}{P_i \cdot q_z} \cdot (p_0 \cdot q_z) = \frac{P_i \cdot q_i}{\left(\frac{P_i \cdot q_z}{P_0 \cdot q_z} \right)} = \frac{P_i \cdot q_i}{I_{Li}} = \frac{C_i}{I_{Li}} = \frac{Y_i}{I_{Li}}$$

donde I_{Li} es un índice de precios Laspeyres con ponderadores fijos en la canasta de consumo q_z .

En resumen, en ambos métodos la utilidad puede ser aproximada por el gasto en consumo o el ingreso, que en este modelo son equivalentes. La diferencia entre enfoques radica en la forma de deflactar las variables monetarias: por un índice de precios tipo Paasche en el caso del enfoque de utilidad monetaria, y tipo Laspeyres en el caso del enfoque de los cocientes de bienestar. En la práctica los índices Laspeyres son más sencillos de computar y están más fácilmente disponibles que los índices Paasche específicos.

Naturalmente, el mundo real es más complicado de lo que este modelo puede ilustrar. Las personas también derivan utilidad de bienes que no compran en el mercado, y por los que no efectúan un gasto, como por ejemplo el consumo de bienes producidos domésticamente, lo cual es habitual en especial en las áreas rurales. En general, esta complicación puede ser acomodada en el marco analítico anterior, siempre que existan precios de referencia para los bienes consumidos no comprados.

El problema se torna más complicado para “bienes” que generan utilidad, pero sin un mercado claramente definido. La libertad, la seguridad, el amor y la salud son determinantes cruciales de la felicidad de las personas, pero difíciles de enmarcar en el modelo anterior. Medir y comparar sólo ingresos o gastos de consumo deja fuera del análisis todos estos importantes aspectos del nivel de vida de las personas. Esta carencia es bien reconocida por la Economía, y de hecho hay varias líneas de investigación que tratan de extender el paradigma de medición monetaria hacia otros bienes y variables. Sin embargo, el problema es complejo y estas extensiones suelen no ser operativas, por lo que la gran mayoría de los estudios distributivos continúan usando el ingreso o el consumo monetario como variables de análisis, con el argumento de que son las aproximaciones al nivel de vida individual más fácilmente mensurables y comparables.

2.3. Ingreso y consumo

El ingreso y el consumo son las dos típicas variables monetarias usadas para aproximar el nivel de vida o bienestar de las personas. Dado que se trata de variables de *flujo*, deben ser definidas en función de un período de referencia: un día, un mes, un año, toda la vida. ¿Cuál es el período de referencia que resulta más relevante para los estudios de pobreza o desigualdad? Por un lado, no es conveniente concentrarse en períodos cortos sujetos a demasiadas fluctuaciones coyunturales. El bienestar en un día particular puede estar muy alejado de los valores habituales para la persona. Nos interesa considerar

algún valor representativo del nivel de vida durante un período suficientemente largo como para aliviar el impacto de shocks circunstanciales de corta duración. Sin embargo, tampoco es conveniente considerar períodos muy largos que impliquen ejercicios artificiales de promediar variables en distintos momentos alejados de la vida. Si el bienestar promedio de una persona a lo largo de toda su vida supera un umbral de pobreza pero la persona sufre privaciones materiales en el 70% de sus años, clasificarla como no-pobre es, cuanto menos, una simplificación excesiva de la realidad.

La literatura distributiva parece converger hacia la elección de un año como período de referencia razonable para medir y comparar los niveles de vida de las personas. Un año es obviamente una medida arbitraria, pero es comúnmente aceptado como un compromiso entre períodos muy cortos (día, semana, mes) sujetos a demasiadas fluctuaciones coyunturales y períodos largos (décadas, toda la vida) en los que se pierde la posibilidad de evaluar el bienestar corriente de las personas. La elección de un período anual como base de referencia también está influenciada por las estacionalidades propias de algunas actividades económicas, particularmente las agropecuarias.

Supongamos entonces que acordamos en comparar el nivel de vida de las personas en un año y que podemos medir el ingreso y el consumo con base anual. ¿Cuál de estas dos variables conviene usar? En el modelo sencillo anterior asumimos que la persona vive un solo período, por lo que es natural que no ahorre, por lo que su gasto en consumo $c \equiv p \cdot q$ es igual a su ingreso y . En la vida real, en cambio, las personas pueden postergar o adelantar consumo mediante decisiones de ahorro o desahorro. Los individuos tienden a ahorrar cuando sus ingresos son excepcionalmente altos de acuerdo a lo previsto, y desahorrar en los momentos en los que los ingresos caen debajo de lo normal. Si la persona vive T años, el problema de optimización que enfrenta es

$$(2.10) \quad \text{Max}_{q_1, \dots, q_T} u(q_1, \dots, q_T) \text{ sujeto a } y_t + r_t s_{t-1} = p_t \cdot q_t + (s_t - s_{t-1}) \quad \forall t = 1, \dots, T$$

donde s_t es el nivel de activos en el año t producto de sus ahorros y r la tasa de interés. La restricción presupuestaria de cada año t indica que el ingreso corriente (igual al ingreso autónomo y sumado al ingreso de capital $r_t s_{t-1}$) es igual al consumo corriente más el incremento o reducción del nivel de activos en ese año.

El supuesto combinado de que los ingresos autónomos (típicamente ingresos laborales y transferencias) son fluctuantes en el tiempo y que las personas tienen preferencias por mantener estables sus niveles de consumo (función u cuasicóncava) introduce una brecha entre el ingreso y el gasto en consumo de cada período. Si suponemos una función aditiva separable $u(q_1, \dots, q_T) = u(q_1) + \dots + u(q_T)$, en cada año la utilidad dependerá sólo del vector de cantidades consumidas en ese año, que valorizadas a los precios de mercado constituyen el gasto anual en consumo $p_t \cdot q_t$. El consumo anual monetario tendría entonces una estrecha relación con el nivel de vida o bienestar individual en cada período. El ingreso monetario anual, en cambio, tendería a sobrestimar el nivel de vida en períodos de ahorro y subestimarlos en años de desahorro.

Una de las principales críticas al uso del ingreso como variable de bienestar reside en la captación del nivel de vida de los adultos mayores retirados del mercado laboral. Algunas de estas personas con escasos o nulos ingresos autónomos mantienen niveles de vida más altos sobre la base de liquidación de activos y desahorros. Estas personas aparecerían como pobres de acuerdo al ingreso corriente, a pesar de gozar de niveles de bienestar más elevados. La sección 4 extiende esta discusión.

Frente a estos argumentos a favor del uso del consumo anual como *proxy* de bienestar individual, algunos esgrimen que el ingreso anual representa las posibilidades genuinas de consumo durante ese período, que la persona es libre de aprovechar, o bien postergar hacia el futuro mediante decisiones de ahorro. En el lenguaje de la sección 2.1, el ingreso anual representaría las *oportunidades* de consumo. Bajo este argumento, si nos interesa el bienestar anual el ingreso anual sería una buena *proxy* en cuanto refleja el acceso a recursos durante ese período, con independencia de lo que se decida hacer con ellos (consumirlos o ahorrarlos).

Al debate de la elección consumo/ingreso se agrega un problema práctico que hasta ahora hemos ignorado: los países no suelen tener encuestas que releven el ingreso o consumo con base anual. Las personas tienen dificultades en recordar información de muchos meses atrás, por lo que las encuestas anuales se vuelven poco confiables. Por otra parte, resulta muy costoso implementar encuestas de panel que entrevisten a las mismas personas mes a mes. De hecho, las encuestas de hogares latinoamericanas suelen preguntar sólo el ingreso del mes previo a la entrevista, y no encuestan reiteradamente en el tiempo a una misma persona. El hecho de que el período de recolección de datos sea corto implica en general un argumento adicional a favor del consumo corriente (mensual) como *proxy* de bienestar en lugar del ingreso corriente (mensual), dado que éste puede estar más afectado por fluctuaciones de corto plazo. El siguiente ejemplo intenta clarificar este punto.

Supongamos dos personas, A y B, con ingresos volátiles a lo largo de los 12 meses del año. El cuadro 2.1 y la figura 2.1 muestran los ingresos y consumos mensuales de estas dos personas en un año típico que, asumamos, se repite sin cambios a lo largo del tiempo.

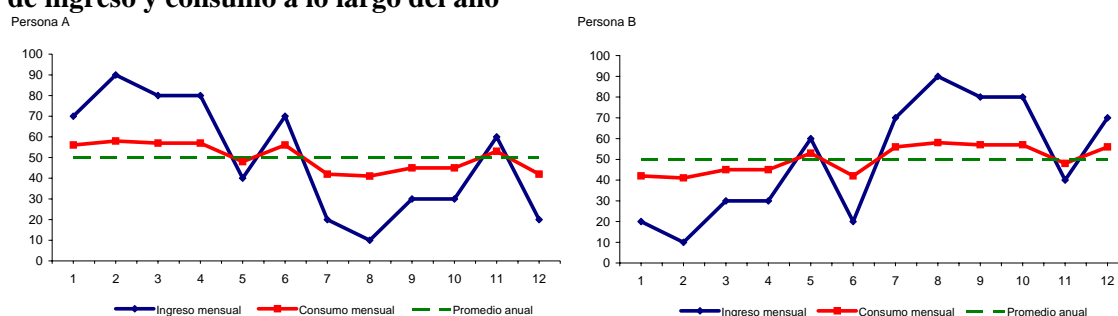
Cuadro 2.1
Ejemplo de dos personas con diferencias en el patrón temporal
de ingreso y consumo a lo largo del año

	Persona A		Persona B		Cociente mayor/menor	
	Ingreso	Consumo	Ingreso	Consumo	Ingreso	Consumo
Enero	70	56	20	42	3.5	1.3
Febrero	90	58	10	41	9.0	1.4
Marzo	80	57	30	45	2.7	1.3
Abril	80	57	30	45	2.7	1.3
Mayo	40	48	60	53	1.5	1.1
Junio	70	56	20	42	3.5	1.3
Julio	20	42	70	56	3.5	1.3
Agosto	10	41	90	58	9.0	1.4
Septiembre	30	45	80	57	2.7	1.3
Octubre	30	45	80	57	2.7	1.3
Noviembre	60	53	40	48	1.5	1.1
Diciembre	20	42	70	56	3.5	1.3
Promedio	50	50	50	50	3.8	1.3

El ejemplo fue construido de modo tal que la única diferencia entre los individuos A y B reside en el patrón temporal de sus ingresos y gastos. El mes de enero para la persona A tiene exactamente los mismos valores de ingreso y consumo que julio para B; febrero de A es igual a agosto de B y así sucesivamente. Obviamente, el ingreso y el consumo promedio durante el año son idénticos para las dos personas (50). Para la persona A los ingresos son claramente superiores durante el primer semestre (lo contrario ocurre con B), aunque también existen fluctuaciones de corto plazo importantes dentro de cada período.

Gráfico 2.1

Ejemplo de dos personas con diferencias en el patrón temporal de ingreso y consumo a lo largo del año



Consistente con la evidencia empírica existente (ej. Deaton, 1992), en el ejemplo las personas “suavizan” su patrón de consumo a lo largo del año.⁷ El ejemplo supone que lo hacen parcialmente, ya sea por ansiedad o urgencias por consumir cuando se percibe el ingreso, o por dificultades para transferir poder adquisitivo de un mes a otro. Se asume adicionalmente que cada año se cierra con la igualdad entre ingresos y consumos, lo cual simplifica la realidad, en la que las personas pueden ahorrar o desahorrar durante períodos de tiempo más largos.

Naturalmente, si estamos interesados en el nivel de vida anual, lo ideal es relevar el ingreso o el consumo anual, que en este caso coinciden para las dos personas. Sin embargo, como explicamos, por problemas prácticos esta opción generalmente no está disponible. Supongamos entonces que el país en el que viven A y B realiza una encuesta en la que sólo se preguntan los ingresos y consumos del mes previo al de la entrevista. Como sugieren el cuadro y gráfico anteriores la “foto” obtenida con una encuesta de este tipo resulta diferente ya sea se tome con la lente del ingreso, o con la del consumo. Si, por ejemplo la encuesta pregunta sobre el mes de enero, reflejará una brecha de ingresos muy grande (70 de A frente a 20 de B), y una brecha de consumo moderada (56 de A y 42 de B). A su vez, cualquiera de estas fotos mensuales es diferente a la obtenida con una lente más amplia con la que se capte al año entero. En esta última, de

⁷ “Suavizar” el consumo hace referencia a mantenerlo en niveles más o menos estables, a pesar que el ingreso se comporte de manera volátil.

hecho, no hay diferencias entre A y B ni en el ingreso, ni el consumo: el promedio anual es igual para ambos.⁸

Nótese que si el objetivo es aproximar el nivel anual de las dos variables, el consumo mensual es una mejor medida que el ingreso mensual ya que no está tan afectado por los vaivenes de la actividad económica. En el ejemplo, en cada mes el valor del gasto de consumo mensual se acerca más al promedio anual que el ingreso mensual.⁹

El último panel del cuadro 2.1 reafirma las ventajas del uso del consumo. La primera columna de ese panel muestra el cociente entre el ingreso mensual de la persona de mayor ingreso y la de menor ingreso. Este cociente es un indicador de desigualdad. Nótese que la posición de persona “más rica” va cambiando entre A y B a lo largo del año. La última columna repite el ejercicio con el consumo. La desigualdad registrada con el consumo mensual es claramente inferior a la registrada con el ingreso mensual y más cercana a la real (si la unidad de tiempo de referencia elegida es el año). Utilizar el consumo mensual implica en el ejemplo una sobrestimación de la desigualdad anual (que es nula), pero el sesgo es inferior al que surge de comparar los ingresos mensuales. Mientras que en promedio el cociente de consumos mensuales entre las dos personas es 1.3, el cociente de ingresos mensuales es 3.8. Naturalmente, cuanto mayor sea el suavizado del patrón de consumo que realizan los individuos, mayores serán las ventajas de utilizar esta variable en lugar del ingreso mensual.

El ejemplo del cuadro 2.1 supone un patrón de ingresos estacional y volátil, y un perfil temporal de consumo más suave. Si bien éste es el caso relevante para muchas personas, no necesariamente lo es para todas. Por ejemplo, los trabajadores asalariados en el sector formal urbano suelen tener salarios fijos, estables en el tiempo, no sujetos a estacionalidades ni shocks de corto plazo. Por su parte el patrón de consumo corriente de ciertas familias puede presentar fuertes estacionalidades y estar sujeto a shocks de corto plazo, dependiendo de preferencias, necesidades y condiciones de mercado. En estos casos, los roles del ingreso y consumo del ejemplo anterior pueden invertirse.

En la evidencia para los países en desarrollo, incluida América Latina, la volatilidad del ingreso es mayor que la del consumo. El cuadro 2.2 muestra el ratio de ingresos entre quintiles extremos, un indicador extendido de desigualdad a explicar con más detalle en el capítulo 6, para las distribuciones del consumo y el ingreso per cápita familiar en algunos países de América Latina. En todos los casos la desigualdad en la distribución del consumo, medida por ese indicador, resulta inferior a la del ingreso.

Cuadro 2.2

Cociente entre quintiles extremos

Distribuciones del consumo y el ingreso per cápita familiar

⁸ Por simplicidad, ignoremos la posibilidad de que difieran si se introducen tasas de descuento.

⁹ Por ejemplo, la suma de las diferencias cuadráticas entre el valor mensual y el promedio anual es 17 veces más grande en el caso del ingreso que del consumo.

Países	año	Quintil5 / Quintil1	
		consumo	ingreso
Ecuador	2006	14.8	20.0
México	2006	12.3	13.4
Nicaragua	2005	7.5	14.8
Panamá	2003	12.9	27.6
Perú	2006	9.0	13.7

Fuente: elaboración propia en base a microdatos de encuestas de hogares.

Si el ingreso reportado en las encuestas está muy “contaminado” por fluctuaciones de corto plazo, en comparación con el consumo, su relación con otros indicadores de privaciones resultará más débil. Algunas personas que son clasificadas como pobres de acuerdo al ingreso reportado en el mes de la entrevista, no tienen privaciones en términos de vivienda, educación y otras variables, ya que sus ingresos de largo plazo (y sus consumos corrientes) son más altos. El gradiente del acceso a ciertos bienes y servicios es menos pronunciado al ser medido con el ingreso que con el consumo.¹⁰ El cuadro 2.3 muestra evidencia en ese sentido para cuatro países latinoamericanos en tres variables: el acceso a una vivienda con materiales no precarios, la disponibilidad de un baño higiénico con descarga de agua y el acceso a un nivel educativo superior.¹¹

Cuadro 2.3

Promedio de variables por quintiles de consumo e ingreso per cápita familiar

México 2006							Nicaragua 2005						
	1	2	3	4	5	Total		1	2	3	4	5	Total
Vivienda de materiales no precarios							Vivienda de materiales no precarios						
Por consumo	35.1	56.8	68.1	77.1	86.3	67.7	Por consumo	65.6	79.5	87.1	92.2	96.3	86.3
Por ingreso	42.1	58.6	65.6	76.2	86.3	67.7	Por ingreso	71.1	80.0	87.5	91.4	94.8	86.3
Disponibilidad de baño y saneamiento							Disponibilidad de baño y saneamiento						
Por consumo	25.6	46.3	64.2	77.4	90.7	65.0	Por consumo	2.3	8.7	17.7	34.9	59.4	29.3
Por ingreso	33.2	48.6	63.1	75.2	90.6	65.0	Por ingreso	3.9	14.0	20.1	35.1	56.4	29.3
Nivel educativo superior (mayores 25 años)							Nivel educativo superior (mayores 25 años)						
Por consumo	0.7	2.2	4.4	10.2	33.8	11.8	Por consumo	0.1	0.6	2.1	4.6	19.5	6.6
Por ingreso	1.4	1.9	4.3	8.7	34.0	11.8	Por ingreso	0.2	0.8	2.1	3.8	20.1	6.6
Panamá 2003							Perú 2006						
	1	2	3	4	5	Total		1	2	3	4	5	Total
Vivienda de materiales no precarios							Vivienda de materiales no precarios						
Por consumo	60.4	85.9	93.8	96.9	99.1	90.9	Por consumo	78.7	76.9	82.6	88.0	94.3	85.0
Por ingreso	71.0	86.5	92.3	96.4	98.2	90.9	Por ingreso	77.8	79.0	82.2	87.5	94.0	85.0
Disponibilidad de baño y saneamiento							Disponibilidad de baño y saneamiento						
Por consumo	11.9	29.9	51.1	70.0	88.7	58.4	Por consumo	32.8	51.6	69.3	81.9	92.7	68.8
Por ingreso	22.1	36.4	54.6	66.3	86.1	58.4	Por ingreso	35.6	53.9	68.5	80.3	91.5	68.8
Nivel educativo superior (mayores 25 años)							Nivel educativo superior (mayores 25 años)						
Por consumo	0.1	1.0	2.9	9.1	30.3	11.0	Por consumo	1.2	4.4	10.8	19.7	39.1	17.4
Por ingreso	0.5	0.7	3.4	8.3	30.4	11.0	Por ingreso	1.6	4.4	9.6	18.8	39.2	17.4

Fuente: elaboración propia en base a microdatos de encuestas de hogares.

Nota: Porcentaje de hogares en viviendas construidas con materiales no precarios; porcentaje de hogares con baño con descarga de agua en la vivienda; y porcentaje de personas mayores de 25 años con estudios superiores. Las definiciones de vivienda de material precario y acceso a baños con saneamiento adecuado difieren entre países. Ver sedlac.econo.unlp.edu.ar.

Existen algunas razones prácticas adicionales a favor del uso del consumo, en lugar del ingreso.¹² Las personas tienden a contestar con más precisión y veracidad sus patrones

¹⁰ El gradiente es la tasa de cambio de una variable a medida que aumenta el ingreso o consumo.

¹¹ La única excepción al mayor gradiente del acceso con respecto al consumo es el caso de la vivienda no precaria en Perú.

¹² Consultar Atkinson (1991), Chaudhuri y Ravallion (1994) y Glewwe y van der Gaag (1990).

de consumo, en tanto tienden a esconder o no recordar sus verdaderas fuentes y niveles de ingreso.¹³ Este problema es particularmente importante en economías primarias y rurales, donde el trabajo por cuentapropia y el consumo de autoproducción son más frecuentes. La captación de actividades ilegales es también más problemática a través del ingreso: las personas tienden a subdeclarar con mayor intensidad los ingresos de ocupaciones ilícitas, que los niveles de consumo sostenidos por el ingreso proveniente de esas actividades.¹⁴

Finalmente, en un estudio distributivo típico nos interesa conocer el nivel de vida alcanzado por la persona *después* de impuestos. Muchas encuestas de los países en desarrollo, entre ellas las latinoamericanas, no suelen incluir preguntas impositivas, por lo que es difícil saber si el ingreso reportado es bruto o neto de impuestos directos. El consumo, en cambio, es un concepto definido después de impuestos directos: uno puede consumir lo que queda del ingreso después del pago de gravámenes.

Dadas estas ventajas, ¿por qué los países no relevan el consumo corriente en lugar (o además) del ingreso corriente? La principal razón es que construir una variable de consumo del hogar no es una tarea sencilla, como veremos más adelante. En parte ello se debe a que es necesario incluir varios cuestionarios tratando de captar cada uno de los bienes y servicios consumidos por la familia, y decidir cuáles de ellos son consumos corrientes estrechamente vinculados al nivel de vida habitual de los miembros del hogar, y cuáles son sólo esporádicos. En la práctica se generan un sinnúmero de situaciones ambiguas para las que no existen recetas homogéneas, por lo que son resueltas de forma dispar por quienes están encargados de construir agregados de consumo, con la consiguiente pérdida de comparabilidad.¹⁵ Estos problemas también ocurren al construir variables de ingreso, pero generalmente son menos graves. Adicionalmente, el trabajo de campo en una encuesta de gastos de consumo es más prolongado y difícil, a menudo requiere varias entrevistas a lo largo del mes, y en consecuencia resulta más oneroso para los gobiernos.

Deaton (2005) argumenta que es más fácil medir el ingreso que el consumo en países ricos, donde la mayor parte de las personas son asalariados formales, y en consecuencia el ingreso es fácil de recordar y reportar, en tanto que las personas son más reacias a cooperar en largas encuestas de consumo que requieren mucho tiempo para contestar. En cambio, medir correctamente ingreso es relativamente más difícil en países pobres, donde muchos individuos están auto-empleados y predominan las actividades agropecuarias con ingresos variables. La magnitud de las fluctuaciones de ingreso y las posibilidades de suavizar consumo mediante el acceso al crédito formal o informal son importantes a la hora de optar por una alternativa de medición.

¹³ Ver Deaton y Grosh (2000) y Meyer y Sullivan (2009).

¹⁴ En contraste, existen consumos en ítems ilegales o que el encuestado prefiere esconder (alcohol, tabaco, drogas, servicios sexuales) aunque los haya abonado con ingresos lícitos.

¹⁵ Ver Beagle *et al.* (2011).

En la práctica, mientras que típicamente los países desarrollados tienen encuestas de hogares sólo con preguntas de ingreso corriente, los países africanos y muchos asiáticos tienen encuestas sólo con preguntas de consumo corriente. América Latina es una región de desarrollo intermedio, donde las ventajas relativas de uno u otro enfoque no son tan claras. En la práctica, en las encuestas de todos los países de la región se releva el ingreso, y en algunas (aproximadamente la mitad) también se incluyen de manera regular cuestionarios destinados a medir el gasto de consumo. La sección 6 trata este punto con más detalle.

2.4. Riqueza

La riqueza es otra variable monetaria de resultado, candidata a ser el centro del análisis distributivo. De hecho, una de las manifestaciones más tangibles de desigualdad es el contraste entre las grandes fortunas y la escasa disponibilidad de activos de vastos sectores de la sociedad. La riqueza de una persona surge de la acumulación de ahorros y de herencias recibidas. A diferencia del ingreso o el consumo, la riqueza es una variable de *stock*, en la que deben incluirse activos inmobiliarios (casas, terrenos, campos), bienes de consumo durables, acciones, bonos y todo otro activo financiero, y restarse las deudas. Existen propuestas de incluir también el stock de capital humano como riqueza, aunque no es práctica usual agregar este ítem.

Conceptualmente, la medición del bienestar a través de la riqueza tiene problemas. Supóngase dos personas, una con ingresos anuales muy altos y otra muy bajos, que no recibieron herencias ni ahorran. La riqueza de ambas personas es cero, pero claramente están en situaciones de bienestar muy diferentes.

Los estudios sobre la distribución de la riqueza pueden ser interesantes *per se* y como contribuciones al estudio de la distribución del bienestar.¹⁶ Desafortunadamente, estos estudios no son usuales en el mundo, y en particular en Latinoamérica, dado que mientras que casi todas las encuestas de hogares relevan el ingreso y/o el consumo, son muy escasas las encuestas con información suficiente para construir una variable de riqueza.

2.5. Ingreso efectivo o potencial

Según discutimos anteriormente, para algunas concepciones de equidad lo relevante no sería el ingreso efectivo de una persona -una variable de resultado-, sino el ingreso potencial que surge de las oportunidades o capacidades con que ese individuo cuenta. En ausencia de diferencias en el ingreso potencial, las diferencias en el ingreso efectivo que surgen de elecciones libres no serían socialmente preocupantes.

El siguiente modelo simple ejemplifica esta distinción. Durante un día una persona dispone de T horas netas para realizar actividades que pueden o no generar ingresos.

¹⁶ Davies y Shorrocks (2000) hacen un exhaustivo análisis teórico de la distribución de la riqueza

Supongamos que toda persona necesita 12 horas para descansar y realizar otras actividades básicas (ej. alimentarse), por lo que el número de horas disponible T es 12. Esas horas pueden ser asignadas a actividades laborales remuneradas l y actividades no remuneradas en el mercado n . Estas últimas actividades suelen ser llamadas en la jerga económica “ocio”, aunque en realidad también abarcan actividades no vinculadas al ocio, como todo trabajo doméstico (limpieza, cuidado de niños, reparaciones en el hogar, etc.). Formalmente, la restricción de tiempo es

$$(2.11) \quad T = l + n$$

El ingreso de una persona es la suma de sus ingresos no provenientes del trabajo M y sus ingresos laborales, que resultan de multiplicar sus horas de trabajo l por el salario horario w . El ingreso en este modelo simple de un período se gasta en el consumo de un bien c con precio unitario.

$$(2.12) \quad M + w.l = c$$

La combinación de las dos ecuaciones anteriores genera la siguiente restricción presupuestaria:

$$(2.13) \quad M + w.T = c + w.n$$

El lado izquierdo de esta ecuación indica el ingreso potencial (*full income*): es el monto que el individuo puede generar, en el caso de usar todo su tiempo disponible (12 horas diarias en el ejemplo) a trabajar en el mercado. Ese ingreso potencial es gastado en consumir bienes y “ocio”, cuyo precio, o costo de oportunidad, es el salario horario de mercado. Algunos investigadores sostienen que lo relevante para un estudio de pobreza y desigualdad es el ingreso potencial, y no el ingreso efectivo $M+w.l$.

Esta alternativa enfrenta al menos dos problemas. La primera es práctica: mientras que el ingreso efectivo es observable, el potencial debe ser estimado. Si la persona decide trabajar T horas en lugar de l su salario horario no necesariamente se mantiene en w , lo cual exige estimar un salario contrafáctico (el que obtendría en una situación – trabajar T horas – que no ha ocurrido en la realidad).

En segundo lugar, el modelo asume que una persona puede voluntariamente elegir el número de horas de trabajo deseadas. A menudo, existen circunstancias en que esto no es posible por situaciones de desempleo involuntario, discriminación u otras restricciones a las elecciones libres en el mercado laboral. Por estas dos razones, en la práctica la gran mayoría de los trabajos distributivos utilizan el ingreso (o consumo) efectivo y no el potencial como variable de análisis.

2.6. Multidimensionalidad

¿Por qué reducir la medición del nivel de vida de una persona a una sola variable? Es creciente el reclamo para que los estudios de pobreza y desigualdad no se limiten a una variable monetaria, sino que capten las distintas dimensiones del bienestar, incluyendo el estado de salud, el acceso a educación y servicios básicos, los niveles de libertad y seguridad, e incluso las percepciones subjetivas y niveles de satisfacción sobre distintos

aspectos de la vida. Si bien conceptualmente la visión multidimensional es comúnmente aceptada, existen dificultades importantes para su implementación práctica. Supongamos dos personas en una sociedad: *A* tiene un ingreso de \$100 y no tiene una fuente de agua potable en la vivienda que habita, mientras que *B* tiene un ingreso de \$80, pero su vivienda está conectada a la red pública de agua. El acceso al agua sin dudas afecta el bienestar de las personas, por lo que es razonable incluir esta variable en un análisis abarcativo de nivel de vida. Pero, ¿cómo hacerlo? La resolución de este problema enfrenta dos dificultades fundamentales. En primer lugar, es muy difícil acordar cuál es la lista precisa de todas las variables que deben ser incluidas como dimensiones del bienestar y cuáles no. En segundo lugar, aun cuando tengamos esa lista, es difícil acordar una estructura de ponderadores que permita pasar de un conjunto de variables a alguna medida resumen. En el ejemplo anterior, ¿*A* goza de un nivel de vida mayor o menor que *B*? Esa evaluación dependerá naturalmente del peso que le demos al acceso al agua. El problema es que no hay maneras objetivas únicas de establecer la estructura de ponderadores. El resultado de esta dificultad es una profusión de indicadores multidimensionales de nivel de vida contruidos con los métodos más variados, y por ende de difícil comparabilidad. Bourguignon (2003), entre otros, reconoce las dificultades para el enfoque multidimensional en alcanzar un grado de operatividad semejante al paradigma unidimensional del ingreso/consumo.

2.7. Bienestar subjetivo

Si el objetivo último de las personas es maximizar su nivel de utilidad o felicidad, ¿por qué no intentar medir y comparar directamente esos niveles? Tradicionalmente la economía ha evitado este camino, por considerar que la utilidad es un concepto difícil de medir y comparar. En contraste con esta tradición, en los últimos años ha florecido la literatura de bienestar subjetivo o “felicidad”, alentada por numerosas encuestas nacionales e internacionales sobre percepciones y niveles de satisfacción. Una pregunta típica de estas encuestas propone al entrevistado ubicarse en algún peldaño de una escalera de acuerdo a su percepción de bienestar individual presente. Las medidas de bienestar subjetivo obtenidas a partir de ésta y otras preguntas similares pueden ser usadas para monitorear el desempeño de una sociedad (como alternativa al PIB per cápita) y también para estudiar desigualdad y pobreza, como veremos en los capítulos 5 y 7.

El papel de las medidas subjetivas de bienestar en la Economía, y en la literatura distributiva en particular, está hoy en pleno debate. Por un lado se subrayan sus potenciales debilidades: las respuestas de las personas a preguntas sobre percepciones pueden estar afectadas por cuestiones culturales o estados de ánimo pasajeros, o basarse en comparaciones muy locales (ej. una persona pobre puede estar satisfecha con su nivel de vida si está rodeada de individuos que sufren mayores privaciones). Un problema adicional potencialmente relevante es el de la *adaptación*: las personas incrementan su utilidad ante una intervención de política positiva (ej. acceso a agua y saneamiento) pero se acostumbran a la mejora muy rápidamente, por lo que el nivel de felicidad puede

volver a valores similares a los pre-intervención.¹⁷ A pesar de estos potenciales problemas, varios autores defienden las mediciones subjetivas como alternativas válidas (y para algunos, superiores) a las mediciones “objetivas” del bienestar, argumentando que conceptualmente son más apropiadas, que generan resultados consistentes e interesantes y que los problemas de medición no son muy diferentes a los de otras alternativas metodológicas.¹⁸ De cualquier forma, el lugar de las mediciones subjetivas es aun secundario, frente al mayor desarrollo y aceptación de las medidas monetarias.

En síntesis, la enorme mayoría de los estudios y estadísticas sobre pobreza y desigualdad se construyen sobre la base de la distribución de variables monetarias que captan el nivel de vida efectivo de las personas, como el ingreso o el consumo. Esta práctica responde en gran medida a las dificultades conceptuales y operativas que enfrentan otras alternativas. Dadas estas dificultades y la previsible correlación positiva entre carencias monetarias y privaciones en otras dimensiones, el enfoque monetario de la medición de la pobreza y la desigualdad ha prevalecido en la práctica sobre alternativas más ambiciosas. Este estado de cosas no implica abandonar el intento por avanzar hacia mediciones subjetivas o multidimensionales del bienestar o de las oportunidades. Por el contrario, la investigación en estas áreas es muy activa y creciente. Pero la complejidad del problema nos exige comenzar el estudio de los temas distributivos restringiéndonos a mediciones unidimensionales monetarias del bienestar. Ese es precisamente el camino que seguimos en el libro, derivando el análisis de otras dimensiones a los capítulos 5 y 7.

3. Personas y hogares

La unidad de análisis en un típico estudio distributivo es la persona. Nos interesa el nivel de vida de las personas, sus carencias y sus diferencias con otros individuos.¹⁹ Ahora bien, las personas no viven aisladas, sino agrupadas en hogares multipersonales que comparten un mismo presupuesto, por lo que el nivel de vida de un individuo está menos vinculado a su ingreso particular, que al ingreso total del hogar y a la forma como éste se reparte entre sus miembros. El hecho que los individuos compartan recursos económicos dentro de unidades familiares introduce un conjunto de interesantes complicaciones en el análisis distributivo.

¹⁷ La famosa *paradoja de Easterlin* indica que los niveles de felicidad no son superiores en países más desarrollados. Esta paradoja ha sido desafiada por nueva evidencia empírica (Deaton, 2008; Stevenson y Wolfers, 2008). Graham y Lora (2009) encuentran que en los procesos de rápido crecimiento la felicidad agregada *disminuye*. El problema de la adaptación es presentado y analizado en Graham (2010) y Gasparini (2010).

¹⁸ Ver Deaton (2008), Layard (2005) y Stevenson y Wolfers (2008). Graham y Lora (2009) es un extenso estudio sobre bienestar subjetivo en América Latina.

¹⁹ Hasta hace algunas décadas la unidad de análisis de los estudios distributivos era el hogar, en lugar del individuo. Esta práctica fue cambiando con el tiempo.

3.1. Hogares

La proporción de personas en hogares unipersonales es minoritaria en América Latina. Un hogar latinoamericano típico es multipersonal (cuadro 3.1). Las diferencias entre países son grandes: mientras que en Uruguay el 79.4% de los hogares son multipersonales, esa proporción asciende a 94.7% en Honduras.

Cuadro 3.1
Proporción de hogares unipersonales y multipersonales

	Honduras	México	R. Dominicana	Uruguay
Hogares Unipersonales	5.3	9.2	12.3	20.6
Hogares Multipersonales	94.7	90.8	87.7	79.4
2 personas	10.1	15.3	16.5	26.0
3 personas	15.6	18.7	19.7	20.6
4 personas	18.9	22.5	20.6	17.4
5 personas	17.9	17.0	15.7	8.7
6 o más personas	32.1	17.3	15.3	6.7

Fuente: elaboración propia en base a microdatos de encuestas de hogares, año 2007.

Pero, ¿qué se entiende por hogar? En general se considera como tal a un grupo de individuos que (i) viven juntos, (ii) consumen bienes y servicios de forma conjunta (principalmente alimentos y vivienda), y (iii) comparten sus ingresos, es decir consumen con cargo al mismo presupuesto común. Si bien a menudo los hogares se conforman con personas que comparten relaciones de parentesco, éste no es necesariamente el caso. Por costumbre, a menudo se utiliza la palabra “familia” o “familiar” (ej. ingreso total familiar), aunque el concepto relevante en el análisis distributivo es el hogar.

En general, cada hogar ocupa una vivienda, aunque se trata de dos nociones diferentes, y en consecuencia la relación no siempre es unívoca. Un hogar puede utilizar más de una vivienda (ej. casas de fin de semana), mientras que una vivienda puede ser ocupada por dos o más hogares. Un ejemplo es el de familias de bajos ingresos que habitan en cuartos de una misma vivienda, que comparten con otras familias. Otro caso usual en muchos países latinoamericanos es el de empleadas/os domésticas/os que viven en la casa de sus empleadores, pero no consumen con cargo al mismo presupuesto.²⁰

3.2. Ajustando por factores demográficos

Dado que las personas viven en hogares donde se comparte el presupuesto, el nivel de vida de una persona está asociado al ingreso (o consumo) total del hogar y a la conformación de ese hogar. En términos analíticos, el nivel de vida x de una persona i perteneciente a la familia h puede escribirse como una función del ingreso total del

²⁰ En la práctica las ambigüedades abundan, problema que es exacerbado por la escasez de información en las encuestas sobre las relaciones dentro del hogar. Dos ejemplos en la larga lista de casos ambiguos son el de empleadas/os domésticas/os que comparten gran parte de las actividades y consumos familiares, aunque reservan una pequeña parte de sus ingresos para gastos propios; y el de adultos jóvenes con pareja e hijos que viven en la casa de sus padres, compartiendo sólo parte de los ingresos y consumos.

hogar Y_h y un vector o conjunto de características demográficas A_h , a discutir en breve.²¹

$$(3.1) \quad x_{ih} = f(Y_h, A_h) \quad \forall i \in h$$

Nótese que ninguno de los argumentos de la función f tiene subíndice i , lo cual es consecuencia de un supuesto fuerte, común en esta disciplina: todas las personas de un hogar comparten exactamente el mismo nivel de vida, lo que implica una distribución *intrahogar* perfectamente igualitaria. La sección 3.4 discute este punto.

¿Cuáles son las características demográficas relevantes A_h que debemos considerar? La más obvia es el número de miembros del hogar. La introducción de esta variable da origen al indicador más extensamente usado de bienestar individual: el promedio simple de los ingresos obtenidos en el hogar, o *ingreso per cápita familiar*

$$(3.2) \quad x_{ih} = \frac{Y_h}{\sum_{j \in h} 1} = \frac{Y_h}{N_h} \quad \forall i \in h$$

Nótese que cada integrante del hogar entra en la ecuación con el mismo peso. Lo único relevante es el tamaño del hogar y no su conformación. Este indicador sencillo es, de hecho, la medida de bienestar individual más utilizada para el estudio de la pobreza y la desigualdad en América Latina. Casi todos los gobiernos de la región usan al ingreso (o consumo) per cápita familiar como indicador de nivel de vida, y computan pobreza oficial en función de esta variable. Investigadores, centros de estudio y organismos internacionales siguen también frecuentemente esta práctica. El monitoreo de la pobreza para los Objetivos de Desarrollo del Milenio de las Naciones Unidas se realiza en función del consumo o ingreso per cápita familiar de las personas.

Ahora bien, en la realidad las personas difieren en ciertas variables que determinan sus necesidades y los recursos que se requieren para cubrirlas. Por ejemplo, el gasto mínimo requerido para satisfacer las necesidades básicas alimentarias es función de ciertas variables como la edad, el género o el nivel de actividad física. A igualdad de ingreso total familiar, no es igual el grado de satisfacción de las necesidades alimentarias al que pueden acceder una pareja de adultos jóvenes, que un hogar compuesto por un adulto y un niño menor, con requerimientos calóricos inferiores.

Estas consideraciones han dado lugar a la creación de *escalas de adulto equivalente*. Se trata de tablas que indican el ratio entre las “necesidades” de cada persona y las de un adulto tipo. Así, si un niño requiriera la mitad de calorías que un adulto, el niño equivaldría a 0.5 de un adulto (medio adulto). Aceptemos la existencia de estas tablas de equivalencia, posponiendo su discusión y estimación hasta la sección 3.3.²²

²¹ La explicación que sigue se desarrolla en términos del ingreso, pero se aplica igualmente al caso del consumo.

²² En UNSD (2005) se reporta que 35% de los países en desarrollo realizan ajustes por escalas de equivalencia en sus mediciones oficiales de pobreza.

Si en lugar de considerar uniformemente a cada integrante del hogar, lo incluimos según su valor de adulto equivalente, obtenemos una nueva medida de nivel de vida individual: el *ingreso familiar por adulto equivalente*

$$(3.3) \quad x_{ih} = \frac{Y_h}{\sum_{j \in h} a_{jh}} \quad \forall i \in h$$

donde a_{jh} es el valor equivalente a un adulto del integrante j perteneciente al hogar h .

Una consideración demográfica adicional surge de la existencia de economías de escala internas al hogar, las cuales permiten que dos personas viviendo juntas mantengan con el mismo ingreso familiar un nivel de vida superior al que es posible en caso de vivir separadas. Estas economías surgen en una variedad de circunstancias en las que es posible reducir el gasto per cápita al compartir el consumo de un bien o servicio. Un caso extremo es el de los bienes públicos del hogar, es decir bienes en los que no hay rivalidad: el consumo de una persona no reduce el consumo por parte de otra. Un cuadro, por ejemplo, es un bien público en el hogar. Si gustan del arte, dos personas viviendo separadas tendrán un cuadro cada una. Si deciden vivir juntas será necesario un solo cuadro, ya que el consumo de los servicios del cuadro (el placer de mirarlo) no se reduce porque otra persona lo consuma simultáneamente. El cargo fijo por vivienda de la conexión a un servicio (agua, electricidad, teléfono) también opera como un bien público que genera economías de escala: la adición de un nuevo miembro al hogar no modifica en nada ese cargo y reduce el costo por persona.

Sin llegar al extremo de los bienes públicos, los hogares explotan los beneficios del consumo conjunto de una amplia variedad de bienes y servicios: el televisor (cuando acuerdan ver el mismo programa), la vestimenta (si se la prestan entre hermanos), el medio de transporte propio (si lo comparten yendo a lugares cercanos), la calefacción, la iluminación y tantos otros. En la alimentación las economías de escala no son tan importantes, dado que se trata de bienes con rivalidad en el consumo: el arroz que come una persona no lo puede comer otra.²³

La siguiente ecuación indica una forma usual de expresar el *ingreso familiar ajustado por economías de escala*

$$(3.4) \quad x_{ih} = \frac{Y_h}{N_h^\theta} \quad \forall i \in h$$

donde θ es un parámetro que puede tomar valores entre 0 y 1. En un extremo, $\theta=1$ indica ausencia de economías de escala. De hecho, en este caso la ecuación de (3.4) se convierte en el ingreso per cápita familiar. En el otro extremo $\theta=0$ indica plenas economías de escala. Este es un caso en el que todos los bienes que se consumen en el hogar son públicos, y por ende el bienestar individual se asocia al ingreso familiar, sin

²³ Sin embargo, aun en la alimentación es posible aprovechar las economías de escala que provienen de la preparación de platos, o del mejor aprovechamiento de la comida sobrante.

ningún ajuste demográfico. Después de todo, en este caso extremo es irrelevante saber cuántas personas hay en el hogar si el consumo de cada miembro no es interferido por la presencia de otros. Naturalmente, en la realidad las economías de escala existen, pero no son tan extremas por lo que θ toma un valor inferior a 1, pero superior a 0. Es importante notar que a mayor valor de θ , *menor* es la intensidad de las economías de escala internas al hogar.

La siguiente fórmula combina los dos ajustes anteriores, expresando el *ingreso familiar ajustado por adulto equivalente y economías de escala*.²⁴

$$(3.5) \quad x_{ih} = \frac{Y_h}{\left(\sum_{j \in h} a_{jh} \right)^\theta} \quad \forall i \in h$$

Una alternativa más sencilla es:²⁵

$$(3.6) \quad x_{ih} = \frac{Y_h}{(M_h + \alpha C_h)^\theta} \quad \forall i \in h$$

donde M_h es el número de adultos en el hogar y C_h es el número de niños. El parámetro $\alpha \in [0,1]$ indica la proporción en la que cada niño equivale a un adulto. Al ingreso ajustado tanto por adulto equivalente como por economías de escala se lo suele llamar *ingreso equivalente*. Los mismos ajustes y definiciones se aplican al caso del consumo.

Una alternativa de cómputo del ingreso equivalente es

$$(3.7) \quad x_{ih} = \frac{Y_h}{1 + \beta(M_h - 1) + \alpha C_h} \quad \forall i \in h$$

De acuerdo a este esquema el primer adulto del hogar es el de referencia, los adultos que se agregan al hogar adicionan β al denominador y los niños α , siendo $\alpha < \beta$.²⁶

Introduzcamos una última consideración conceptual. Algunos analistas sostienen que dado que la decisión de conformación familiar es enteramente libre y racional, y que la utilidad derivada de la presencia de hijos no es mensurable, es recomendable no realizar ningún ajuste demográfico y aproximar el nivel de vida de una persona directamente por su ingreso familiar, es decir $x_{ih} = Y_h$. Supongamos dos parejas con un ingreso familiar fijo de \$1000. La primera decide tener 6 hijos y la segunda 2. Aun considerando los ajustes por adulto equivalente y economías de escala, el nivel de vida estimado de los integrantes de la primera pareja será inferior al de la segunda. Ahora bien, nótese que en

²⁴ Ver, por ejemplo, Buhmann *et al.* (1988).

²⁵ National Research Council (1995) y Deaton (2003).

²⁶ La escala de Oxford (o antigua escala OECD) propone $\beta=0.7$ y $\alpha=0.5$; mientras que la escala modificada de OECD usa $\beta=0.5$ y $\alpha=0.3$. Ver Atkinson *et al.* (1995). La OECD utiliza también una variante en la que el primer adulto del hogar vale 0.67, cada adulto adicional o niño mayor de 14 años agrega 0.33, y cada niño menor de 14 adiciona 0.2. De este modo una pareja suma 1 y constituye el hogar de referencia (en lugar del adulto de referencia), mientras que una familia con dos adultos y dos niños menores suma 1.4 (Cowell, 2008).

el ejemplo esta diferencia surge enteramente de las distintas decisiones de fecundidad. La primera pareja podría haber tenido 2 hijos, pero decidió tener 6, siendo consciente de que esto implicaba una reducción en el nivel de vida, al menos corriente. De acuerdo al paradigma de las oportunidades, algunos sostienen que las dos parejas deberían ser consideradas en igualdad de condiciones, pese a gozar de distintos niveles de acceso a bienes materiales, y en consecuencia la situación desigual descripta arriba no se correspondería con ninguna inequidad social.

En la práctica, esta línea de argumentación en general es desechada por la literatura distributiva en función de dos razones. Primero, las decisiones de fecundidad no siempre son enteramente racionales, ni basadas en el mismo conjunto de información disponible para toda la población. En esta línea, hay quienes argumentan que condicionamientos culturales y restricciones informativas son responsables de las mayores tasas de fecundidad en los estratos sociales de menores ingresos. En segundo lugar, si bien las parejas pueden decidir libre y racionalmente el número de hijos, éstos no pueden elegir el tamaño del hogar en el que nacen y se crían. El tamaño de la familia es una variable de elección para los padres, pero es una variable de circunstancia (exógena) para sus hijos. Un niño no elige nacer en un hogar con carencia de recursos materiales per cápita producto del mayor número de integrantes. Atento a estos argumentos, los factores demográficos deberían ser considerados a la hora de evaluar y comparar los niveles de vida individuales.

El siguiente ejemplo ilustra el impacto que los ajustes demográficos pueden tener sobre los resultados distributivos. Supóngase que existen dos familias cuyos ingresos totales son iguales a \$1000. La familia A está integrada por dos adultos y la familia B por dos adultos y dos niños. El ingreso per cápita familiar es \$500 para cada integrante de A y \$250 para cada integrante de B, por lo que el ratio de ingresos (una medida posible de desigualdad) es 2.

Cuadro 3.2
Cambios en el ratio de ingresos por ajustes demográficos

Familia	Demografía		Ingreso							
			por adulto equivalente			ajuste por economías de escala				
	M	C	total	per cápita	$\alpha=0.75$	$\alpha=0.5$	$\alpha=0.25$	$\theta=0.75$	$\theta=0.5$	$\theta=0.25$
A	2	0	1000	500	500	500	500	595	707	841
B	2	2	1000	250	286	333	400	354	500	707
ratio ingresos			1	2	1.8	1.5	1.3	1.7	1.4	1.2

Nota: M=número de adultos, C=número de niños

Esta brecha se reduce si consideramos el ajuste por adulto equivalente. A partir de la ecuación (3.6), si asumimos ausencia de economías de escala y $\alpha=0.75$, el cociente de ingresos entre los integrantes de las familias A y B se reduce de 2 a 1.8. Cuanto menor es α , menor es la ponderación dada a un niño (menor es su carga sobre el presupuesto familiar), lo cual favorece más a la familia B: con $\alpha=0.5$ el cociente de ingresos es 1.5, y con $\alpha=0.25$ el ratio cae a 1.3.

El último panel del cuadro ilustra el impacto del ajuste por economías de escala, ignorando la escala de adulto equivalente. Cuanto menor es el valor de θ mayor es la intensidad de las economías de consumo internas al hogar, lo cual favorece relativamente más a la familia más numerosa (la B) y en consecuencia reduce la brecha estimada de niveles de vida: con $\theta=0.75$ el cociente es 1.7, mientras que con $\theta=0.25$ el ratio cae a sólo 1.2.

3.3. Los parámetros de ajuste

Hasta acá hemos tomado como un dato la existencia de parámetros de ajuste demográfico, como θ y α , y discutido sobre la conveniencia y el impacto de usarlos en estudios distributivos. Desafortunadamente, en la práctica existen numerosos problemas conceptuales a la hora de definir escalas de adulto equivalente o de estimar economías de escala internas al hogar. De hecho, gran parte de la literatura ha terminado por adoptar parámetros más o menos arbitrarios, o directamente por desechar todo ajuste y trabajar con el ingreso (o consumo) per cápita familiar.

Escalas de equivalencia

Idealmente, una escala de equivalencia debe captar las diferencias en el costo de alcanzar un determinado nivel de vida como consecuencia de las distintas estructuras demográficas del hogar. Analíticamente,

$$(3.8) \quad e_h = \frac{e(A_h, p; u)}{e(A_0, p; u)}$$

donde $e(\cdot)$ es una función que indica el costo mínimo de obtener un nivel de utilidad u , dados los precios de los bienes p , y un vector de características demográficas A (miembros del hogar, edades, etc.). El valor de e_h representa entonces el costo relativo para una familia con características A_h de alcanzar el nivel de vida u respecto de una familia “base” con características A_0 . El ingreso equivalente sería entonces simplemente Y_h/e_h .

En la práctica el problema es que e_h depende del nivel elegido de u y del hogar base A_0 , por lo que existen infinitas posibles escalas. Aun eludiendo estas dificultades, la estimación de las escalas a partir de una ecuación como la (3.8) requiere de muchos supuestos fuertes y no puede hacerse simplemente analizando la estructura de gasto familiar (Deaton y Muellbauer, 1989).

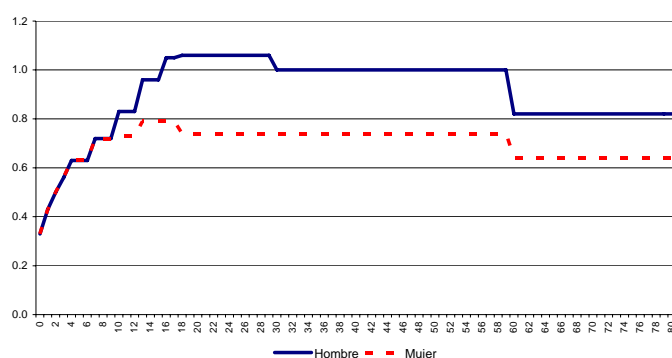
Operativamente, las escalas de equivalencia provienen de tres tipos de estudios: nutricionales, de evaluación subjetiva de necesidades y de comportamiento. A menudo, sin embargo, la insatisfacción con los problemas metodológicos de los tres enfoques lleva a que investigadores y oficinas de estadística utilicen escalas *ad hoc*.²⁷

²⁷ Deaton (1997) es una excelente referencia para iniciarse en la literatura sobre escalas de equivalencia.

Escalas nutricionales

Algunas escalas provienen de estudios nutricionales que comparan el costo estimado de una canasta básica de alimentos que permite satisfacer un determinado nivel de requerimientos calóricos, los cuales dependen de la edad y otras características de la persona. El gráfico siguiente ilustra la escala de adulto equivalente utilizada para los cálculos de pobreza oficial en Argentina basada en un estudio nutricional (Morales, 1988). La persona de referencia es un varón adulto entre 30 y 59 años ($a_j=1$). Los menores de 1 año equivalen a un tercio del hombre adulto de referencia. Ese valor crece con la edad hasta alcanzar 0.72 a los 10 años, edad en la que la escala se separa entre varones y mujeres, reconociendo necesidades calóricas diferenciales por género. El valor de adulto equivalente supera el 1 para los varones entre 16 y 29 años, y se reduce significativamente, tanto para varones como para mujeres, a partir de los 60 años.

Figura 3.1
Escala de adulto equivalente oficial
Argentina



Fuente: INDEC, Argentina.

El problema con las escalas nutricionales es que sólo tienen en cuenta las necesidades básicas *alimentarias* de las personas. Un niño puede necesitar menos alimentos que un adulto, pero sus requerimientos en educación, salud o vestimenta pueden ser superiores, con lo que la escala pierde sentido. La brecha en algunas escalas entre hombres y mujeres, razonable en términos de las necesidades calóricas de ambos géneros, carece de sentido al incluir a la totalidad de los bienes y servicios consumidos.²⁸ Como veremos en el capítulo 4, aun las personas clasificadas como indigentes destinan una parte significativa de su presupuesto al consumo de bienes no alimentarios.

Típicamente, las escalas de equivalencia presentan cortes abruptos que generan resultados controversiales. Tomemos una escala bastante detallada como la usada en Argentina y graficada en la figura 3.1, y consideremos tres personas idénticas salvo por

²⁸ En la llamada escala de Amsterdam el hombre adulto (18 años o más) sirve de referencia, los hombres entre 14 y 17 valen 0.98, las mujeres mayores de 14 0.9, y los niños/as menores de 14 equivalen a 0.52 de un hombre adulto. Estas escalas con diferencias de género han perdido importancia en el tiempo.

diferencias de género, y diferencias muy pequeñas en la edad. *A* es un hombre de 29 años, *B* un hombre de 30 y *C* una mujer de 30, todos viviendo solos, y con un ingreso de \$250 mensuales cada uno. Es difícil argumentar que estas tres personas tienen acceso a niveles de vida significativamente diferentes. Sin embargo, el uso de la escala de equivalencia nos lleva a los ingresos equivalentes de la última columna del cuadro 3.3. De acuerdo a estos valores, el nivel de vida de *B* sería 6% superior al de *A*, mientras que el de la mujer *C* sería 43% superior al de *A*.

Cuadro 3.3

**Ingreso por adulto equivalente de tres personas
sobre la base de la escala de equivalencia oficial en Argentina**

	Sexo	Edad	Adulto equivalente	Ingreso	Ingreso equivalente
A	Hombre	29	1.06	250	236
B	Hombre	30	1.00	250	250
C	Mujer	30	0.74	250	338

El enfoque subjetivo

Una alternativa a las escalas nutricionales proviene de la evaluación individual subjetiva de las necesidades (van Praag y Wamaar, 1997). Este enfoque está basado en las respuestas en encuestas a preguntas del tipo ¿cuánto ingreso cree que es necesario para cubrir las necesidades básicas? Supongamos que llamamos con Y^s a la respuesta a esta pregunta, la cual en la práctica es una función creciente del ingreso efectivo del hogar (Y) y del número de miembros (N), entre otros factores.

$$(3.9) \quad Y^s = f(Y, N) \text{ con } f_Y > 0 \text{ y } f_N > 0$$

donde f_j indica la derivada de f respecto del argumento j . Llamemos con Y^n al nivel de ingreso para el cual el ingreso requerido reportado Y^s coincide con el efectivo Y . Se argumenta, lo cual no está exento de controversias, que ese nivel puede ser interpretado como una medida de “necesidad”.

$$(3.10) \quad Y^n = f(Y^n, N)$$

En la práctica es usual estimar esta relación con un modelo de regresión lineal en logaritmos con datos al nivel de familias h

$$(3.11) \quad \ln Y_h^s = \alpha + \beta \ln N_h + \gamma \ln Y_h + \varepsilon_h$$

Haciendo $Y_h^s = Y_h = Y_h^n$

$$(3.12) \quad \ln Y_h^n = \left(\frac{1}{1 - \gamma} \right) (\alpha + \beta \ln N_h + \varepsilon_h)$$

Nótese que $\beta/(1-\gamma)$ es en este contexto una medida de la elasticidad de las necesidades (Y_h^n) con respecto al tamaño familiar, y por ende estima las economías de escala internas al hogar. De hecho, en los trabajos empíricos las estimaciones de $\beta/(1-\gamma)$ son inferiores a 1, sugiriendo que un aumento proporcional en el tamaño del hogar está asociado a un aumento menos que proporcional de sus “necesidades”. Si bien este enfoque ha sido utilizado en ocasiones, su popularidad es limitada dado que depende de interpretaciones controversiales acerca del valor de Y^n , de la confianza a las respuestas de preguntas ambiguas como la mencionada arriba y de supuestos econométricos necesarios para la estimación e interpretación de los coeficientes.²⁹

Enfoque de comportamiento

El enfoque de comportamiento está basado en el examen de las estructuras de gasto de las familias. El método más conocido es el de Engel, según el cual el costo de un niño adicional puede aproximarse por el gasto incremental de consumo que habría que asignar a una familia para que recupere el valor del *share* en alimentos original (usado como *proxy* inversa del nivel de bienestar), antes del nuevo niño. El método, sin embargo, no tiene en cuenta los cambios genuinos en la estructura de consumo de una familia ante la presencia de un niño, en particular la tendencia a incrementar el consumo de alimentos, lo que lleva a la sobrestimación del costo de un niño. Problemas semejantes se presentan al intentar usar el método para estimar las economías de escala internas al hogar.

Los estudios empíricos basados en datos de gastos de consumo familiar sugieren que la elasticidad del costo de vida con respecto al número de miembros decrece con la importancia de los bienes públicos en el hogar. En un extremo, en un hogar donde todos los bienes consumidos son privados, esa elasticidad es máxima.³⁰ Los resultados de este enfoque, aunque intuitivamente útiles, son insuficientes para identificar una escala de equivalencia precisa, por lo que en la práctica su uso termina siendo muy limitado (Deaton, 1997).

Una alternativa ingeniosa, aunque también limitada, consiste en examinar el gasto en bienes que no son consumidos por los niños. El consumo de “bienes de adultos” debería disminuir ante la presencia de un niño que requiere recursos y no agrega demanda en esos bienes. El costo de un niño de acuerdo a este método, conocido como *Rothbarth*, puede ser aproximado por el monto de compensación necesario para que los padres recuperen el nivel de gasto en bienes de adultos, previo a la existencia del niño.

Escalas ad hoc

²⁹ Nótese que es muy posible que existan factores no observables que afecten tanto a N_h como a Y_h^n .

³⁰ Véase Drèze and Srinivassan (1997) y Deaton (1997).

La imposibilidad de arribar a una escala de equivalencia no controversial ha implicado que la mayoría de los investigadores se incline por soluciones prácticas. Una alternativa generalizada y sencilla es fijar de forma *ad hoc* los parámetros de una ecuación que defina el ingreso equivalente. Como hemos visto, una fórmula usual es $Y_h/(M_h + \alpha C_h)^\theta$ donde α indica el costo diferencial de un niño respecto a un adulto, mientras que θ regula el grado de economías de escala. En países pobres donde los alimentos son un componente muy importante del consumo familiar, el costo de los niños es inferior al de los adultos en función de sus menores necesidades de calorías, y por ende α es un valor relativamente bajo. Varios trabajos han usado valores de α en el entorno de 0.25. En cambio, en economías más desarrolladas donde el consumo se extiende a otros bienes como educación, salud y entretenimiento, el costo de los niños se vuelve semejante al de los adultos y α tiende a 1.

Respecto de las economías de escala, éstas son escasas en economías primarias donde el presupuesto se gasta mayoritariamente en alimentos, es decir en bienes privados en los que hay rivalidad en el consumo. En este caso θ debería fijarse en un valor alto, cercano a 1. En cambio en economías más desarrolladas en las que los alimentos ocupan una fracción menor del presupuesto, y en los que varios bienes y servicios de uso conjunto son importantes, las economías de escala resultan más relevantes, por lo cual θ debería fijarse en un valor menor (0.75 ha sido usado en varios estudios para países desarrollados).

América Latina es una región de nivel de desarrollo medio en la que ninguno de los dos extremos discutidos arriba parece ajustarse, por lo que sugerimos trabajar con niveles intermedios (ej. $\theta=0.9$ y $\alpha=0.7$). En teoría, sería conveniente practicar ajustes diferentes entre países, dado que éstos difieren en sus niveles de desarrollo y por ende en sus estructuras de consumo. El argumento, de hecho, podría ser extendido hasta considerar la posibilidad de diferentes ajustes para distintos tipos de hogares al interior de cada país. De esta discusión es claro que, dada la arbitrariedad en la elección de los parámetros, resulta muy relevante hacer un análisis de sensibilidad de los resultados. En los capítulos 4 y 6 retomamos este punto al evaluar el impacto de los ajustes demográficos sobre la medición de la pobreza y la desigualdad, respectivamente.

3.4. La distribución intrahogar

Prácticamente todos los estudios distributivos hacen implícitamente un supuesto clave: la distribución del bienestar dentro del hogar es perfectamente igualitaria. Esto implica que todos los miembros del hogar gozan del mismo nivel de vida, cualquiera sea la forma como éste se mida y por tanto comparten el mismo status de pobreza: no es posible bajo este supuesto que un miembro del hogar sea pobre, mientras otro no lo sea.

Existe evidencia empírica que contradice el supuesto general de desigualdad intrahogar nula.³¹ En algunos casos existen sesgos en la asignación del presupuesto familiar hacia

³¹ Ver Thomas (1990) como ejemplo de una creciente literatura.

quienes aportan una fracción mayor de los recursos del hogar, o hacia quienes mantienen una posición de poder (frecuentemente los hombres).

De cualquier forma, la dificultad en estimar con precisión el grado de desigualdad intrahogar y en obtener resultados generales hace que, salvo en estudios específicos sobre el tema, el ajuste por este factor sea ignorado en las mediciones distributivas. El impacto de esta práctica sobre la desigualdad es claro: al ignorar diferencias intrahogar la desigualdad es subestimada.³²

Supongamos el siguiente ejercicio hipotético. En cada hogar los recursos se distribuyen de manera sesgada hacia el jefe de hogar (usualmente un hombre) de una forma concreta: hay un “impuesto” sobre el ingreso total del hogar que se asigna enteramente al jefe. El resto del ingreso se reparte igualitariamente entre todos los miembros de la familia. Así, en este ejercicio el nivel de vida de cada persona del hogar puede ser aproximado por

$$(3.13) \quad x_{ih} = \frac{Y_h(1-t)}{N} + s_{ih} \quad \text{donde } s_{ih} = tY_h \text{ si la persona } i \text{ es el jefe y } 0 \text{ para el resto}$$

El siguiente cuadro muestra el cociente de ingresos entre los deciles 10 y 1 (un posible indicador de desigualdad) ante un conjunto de tasas t diferentes. Bajo el supuesto de completa igualdad intrahogar el cociente de ingresos entre deciles extremos en Paraguay es 42.9. Ese indicador de desigualdad aumentaría a 45.5 con un “impuesto” para el jefe de 10%, y a 61.7 si la distribución intrahogar estuviera fuertemente sesgada hacia el jefe. Este ejercicio ilustrativo sugiere que el supuesto de igualdad intrahogar puede no ser inocuo en la práctica.

Cuadro 3.4

Cociente de ingresos entre deciles extremos Ejercicios alternativos asumiendo redistribuciones lineales intrahogar hacia el jefe

tasa	Argentina	Honduras	Paraguay	Venezuela
0.0	34.5	74.2	42.9	27.9
0.1	36.5	78.4	45.5	29.9
0.2	41.2	88.7	51.9	35.0
0.3	48.7	105.3	61.7	43.0

Fuente: elaboración propia en base a microdatos de encuestas de hogares.

3.5. Cálculos por individuos u hogares

Dado el supuesto usual de perfecta igualdad al interior de cada hogar, resulta indiferente realizar cualquier cálculo directamente a nivel de individuos, o alternativamente hacerlo a nivel de hogar, con el cuidado de ponderar por el tamaño del hogar. Supóngase una familia h con N_h miembros. El supuesto de igualdad interna al hogar implica que en

³² El resultado sobre la pobreza es ambiguo. Ver Haddad y Kanbur (1990).

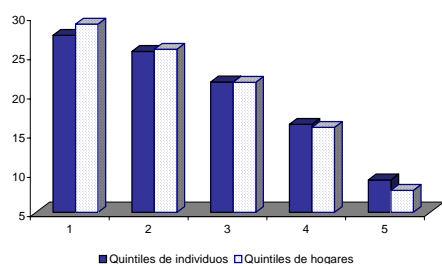
cualquier cálculo $x_i = x_h$ para toda persona i perteneciente a h . En este caso es equivalente incluir x_i una cantidad N_h de veces (análisis por individuos), o incluir x_h ponderado por N_h (análisis por hogares).

Algunos investigadores e institutos de estadística latinoamericanos reportan resultados de pobreza o desigualdad a nivel de hogar sin ponderar por tamaño. Naturalmente esta práctica sí lleva a resultados diferentes. En particular, dado que las familias pobres suelen ser más numerosas, la tasa de pobreza calculada sobre hogares es menor a la calculada sobre individuos (ver capítulo 4).

Es usual presentar la distribución de variables en cuantiles de ingreso o consumo. En este caso, es importante puntualizar si se trata de cuantiles contruidos con igual número de individuos o igual número de hogares. En algunos casos los resultados pueden diferir. La figura 3.2 muestra la participación de cada quintil de la distribución del ingreso per cápita familiar en el total de beneficiarios de la educación pública primaria en Honduras en el año 2004. En el primer caso los quintiles están contruidos sobre la base del total de personas, mientras que en el segundo la unidad de análisis es el hogar (sin ponderar por el número de miembros). El grado de focalización del programa de educación primaria pública es algo mayor en el segundo caso. La razón es que el primer quintil de hogares reúne al 20% de las familias hondureñas, pero dado que se trata de hogares más numerosos, agrupa a un porcentaje mayor de personas (23%), particularmente de niños en edad escolar, que se benefician de la escuela primaria pública. A lo largo del libro, y siguiendo la tendencia de la literatura distributiva, trabajaremos con cuantiles de individuos y no de hogares, salvo que se indique lo contrario.

Figura 3.2

Participación de cada quintil en el total de asistentes a las escuelas primarias públicas Honduras, 2004
Quintiles de ingreso per cápita familiar por individuos y hogares



Fuente: elaboración propia sobre la base de microdatos de la ENCOVI, 2004.

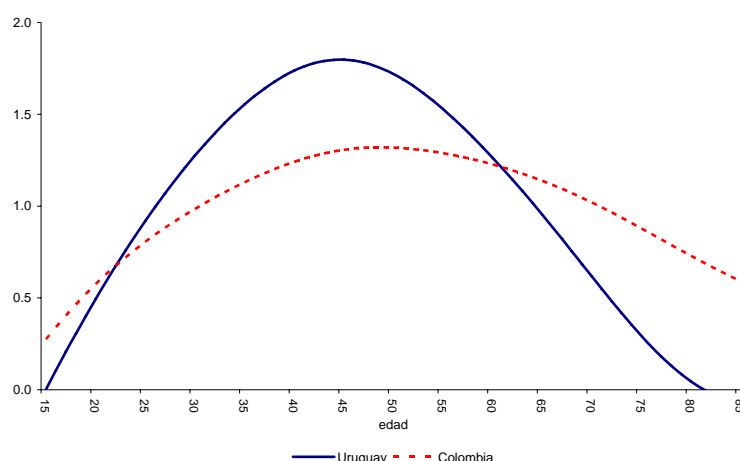
4. Edad y comparaciones intertemporales

Unas de las razones por las cuales los niveles de vida difieren entre personas es la edad. Típicamente existe un “ciclo de vida” por el cual los individuos tienen acceso a diferentes recursos a medida que avanza su vida. La consideración de la edad y el tiempo introduce dificultades adicionales en el análisis distributivo.

4.1. Ciclo de vida

Existe un claro patrón temporal de los ingresos laborales que las personas generan a lo largo de su vida: escasos o nulos en la niñez, crecientes en la juventud, máximos en la edad adulta y decrecientes en la tercera edad. La figura 4.1 ilustra este patrón de ingresos laborales y edad en forma de U invertida para Colombia y Uruguay.

Figura 4.1
Ingreso laboral por edad

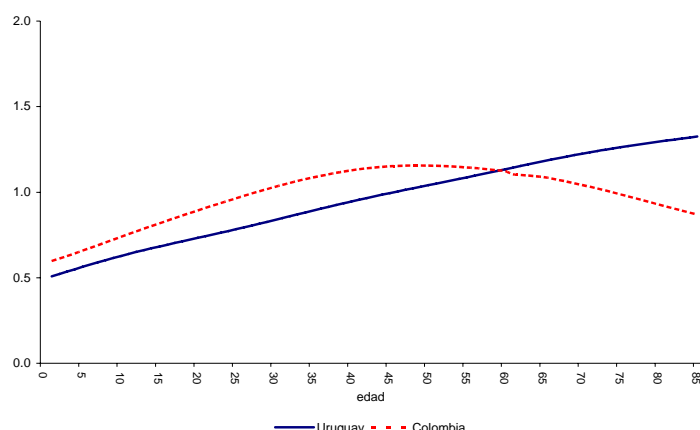


Nota: estimaciones no paramétricas (*lowess*). Series normalizadas a la media de cada país.

Fuente: elaboración propia sobre la base de microdatos de las encuestas continuas de hogares de Colombia, 2006 y Uruguay, 2007.

Nótese que el patrón es más marcado en el caso de Uruguay, un país con mayor desarrollo educativo, y por ende con menor inserción de su juventud en el mercado laboral, y con un sistema de pensiones de mayor cobertura, lo cual explica los bajos ingresos laborales de su población adulta mayor. Es interesante notar que en general este patrón se repite, aunque con menor intensidad, al considerar variables asociadas al nivel de vida de la persona y no a sus ingresos laborales individuales. Típicamente, el ingreso familiar per cápita es menor en los niños, dado que éstos suelen vivir en hogares con más miembros, con padres jóvenes usualmente en el comienzo de su vida laboral, y por ende con salarios más bajos y escaso capital. El nivel de vida suele aumentar hasta la edad adulta, cuando el tamaño de la familia se reduce, el salario alcanza un máximo y se acumula algún capital. El nivel de vida en la vejez depende del desarrollo de los sistemas de seguridad social. En varios países latinoamericanos, con sistemas de pensiones frágiles, la tercera edad es una etapa de mayores privaciones. La figura 4.2 muestra el ingreso per cápita familiar para diferentes edades en Colombia y Uruguay. En el caso de Colombia este indicador de bienestar individual cae considerablemente a partir de la edad adulta; en contraste el extendido sistema de pensiones en Uruguay permite a los adultos mayores de ese país del Cono Sur mantener niveles de vida relativamente altos.

Figura 4.2
Ingreso per cápita familiar por edad



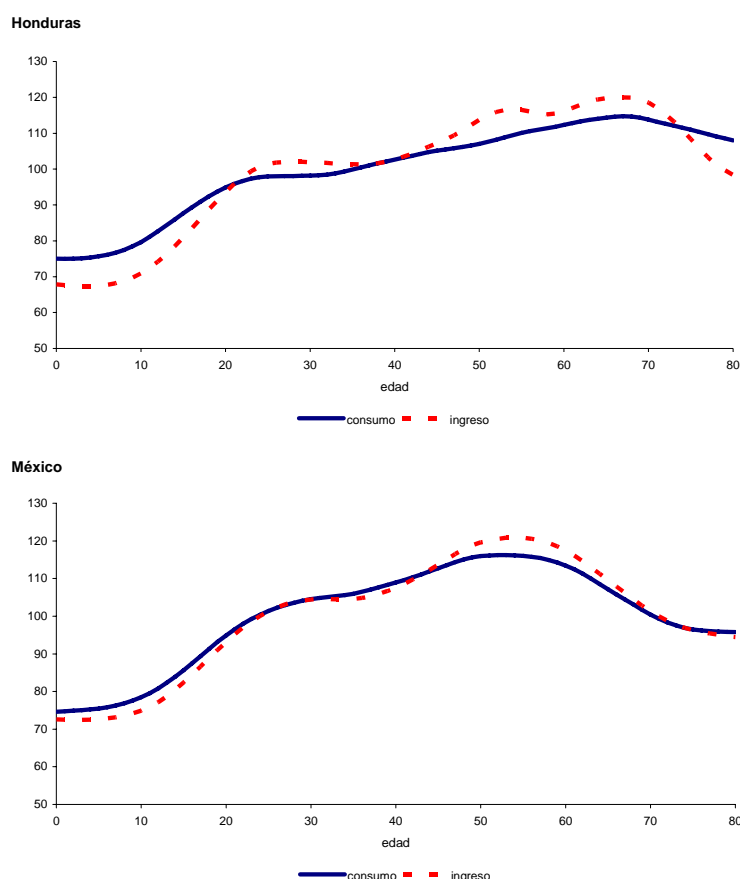
Nota: estimaciones no paramétricas (*lowess*). Series normalizadas a la media de cada país.

Fuente: elaboración propia sobre la base de microdatos de las encuestas continuas de hogares de Colombia, 2006 y Uruguay, 2007.

La consideración del consumo como variable de bienestar no cambia las principales conclusiones, aunque las suaviza. Si las personas tienen acceso al crédito, formal o informal, pueden endeudarse en la juventud, pagar las deudas y ahorrar cuando adultas y desahorrar en la vejez. Este comportamiento les permitiría mantener un patrón de consumo más estable en el tiempo, no tan vinculado a sus ingresos corrientes. La figura 4.3 ilustra el ingreso y el consumo familiar per cápita para diferentes edades en Honduras y México. La forma de U invertida es sólo ligeramente menos notoria en el caso del consumo, lo que genera ciertas dudas acerca de la fuerza de estos argumentos, y debilita en algo la posición de quienes sólo admiten aproximar el nivel de vida con datos de consumo.³³

Figura 4.3
Ingreso y consumo per cápita familiar por edad

³³ La metodología usual consiste en asignar el mismo valor del consumo (ajustado por factores demográficos) a cada integrante del hogar. Existen propuestas alternativas de estimar la real distribución del gasto entre miembros de una familia, argumentando que el consumo de varios bienes y servicios es función de la edad (ej. mayor gasto en educación en jóvenes, salud en adultos mayores, etc.). Bajo esta propuesta el patrón etario del consumo luce más suave que en la figura 4.3. La metodología de esta alternativa puede consultarse en el *National Transfer Accounts Project* www.ntaccounts.org.

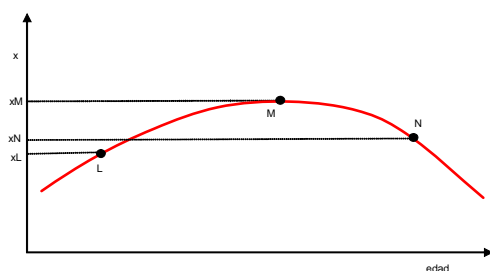


Fuente: elaboración propia sobre la base de microdatos de la ENCOVI 2004 de Honduras y la ENIGH 2006 de México.

4.2. Comparaciones entre edades

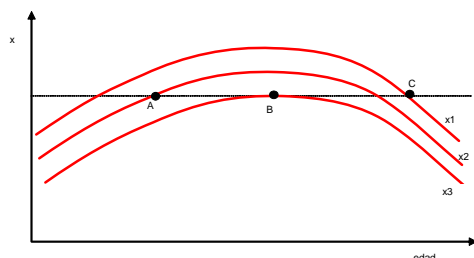
Algunos analistas sostienen que es inválido considerar como inequitativas diferencias de nivel de vida que sólo provienen de edades diferentes. Para ilustrar este punto la figura 4.4 representa el patrón de ingresos de un grupo de personas para las que el nivel de vida x sólo varía por la edad. La persona L es un joven con un nivel de vida relativamente bajo x_L , pero que con el tiempo accederá al nivel de un adulto x_M y luego al de un adulto mayor x_N . Si bien existe desigualdad en un momento del tiempo (la “foto”), no hay diferencias en el nivel de vida intertemporal o permanente de estas personas (“la película”).

Figura 4.4
El ciclo de vida
Desigualdad sobre el ingreso corriente
sobrestima la desigualdad intertemporal



La siguiente figura ilustra un caso en el que mirar la “foto” lleva a subestimar las desigualdades, en lugar de sobrestimarlas, como en el caso anterior. Supóngase que el perfil de ingresos por edad en la región 1 (línea x_1) es superior al de la región 2 (x_2), y éste a su vez mayor al de la región 3 (x_3). La persona B pertenece a la región más desfavorecida, pero se encuentra justo en el punto más alto de su perfil, a diferencia de las personas A en la región 2 y C en la región 1. Un estudio que considere el nivel de vida actual de estas tres personas indicará ausencia de desigualdad, pese a que en promedio a lo largo de sus vidas los tres individuos gozarán de niveles de vida muy diferentes.³⁴

Figura 4.5
El ciclo de vida
Desigualdad sobre el ingreso corriente
subestima la desigualdad intertemporal



En base a este argumento algunos investigadores proponen restringir el análisis distributivo a grupos etarios específicos. En general estos estudios se concentran en el grupo de adultos en edad laboralmente activa (aproximadamente entre 25 y 65 años). Naturalmente, esto implica descartar del análisis una parte importante de la población. El cuadro 4.1 muestra cómo en el caso de algunas distribuciones latinoamericanas, la desigualdad en la distribución del ingreso per cápita familiar (medida con el coeficiente de Gini, un indicador a desarrollar en el capítulo 6) de toda la población no es muy diferente a la de grupos de adultos en edad laboral activa.

³⁴ Naturalmente, el resultado del ejemplo depende de elegir sólo algunas personas en cada región. En la realidad, en cada región existen personas en cada rango etario por lo que el signo del sesgo no es claro.

Cuadro 4.1**Desigualdad en la distribución del ingreso per cápita familiar por grupo etario
Coeficientes de Gini**

	Brasil	El Salvador	Perú
Todos	54.8	49.6	49.7
Rango etario			
[25-64]	54.2	49.8	50.3
[30-39]	54.5	49.6	50.9
[40-49]	53.2	47.6	49.0
[50-59]	54.1	48.5	50.6

Fuente: elaboración propia en base a microdatos de encuestas de hogares, año 2007.

4.3. Nivel de vida intertemporal

Una posibilidad sostenida por algunos analistas es la de comparar los niveles de vida intertemporal o “promedio” durante toda la vida de las personas. Obviemos los problemas introducidos por la formación y destrucción de hogares y supongamos que las personas viven siempre solas. En ese contexto la restricción presupuestaria intertemporal de una persona típica estaría dada por la siguiente ecuación:

$$(4.1) \quad \sum_{t=0}^T \frac{y_t + s_0}{(1+r)^t} = \sum_{t=0}^T \frac{c_t + s_T}{(1+r)^t}$$

donde t indexa al tiempo que transcurre entre que la persona nace (en $t=0$) y muere (en $t=T$), y son ingresos autónomos provenientes del trabajo o transferencias, c es consumo, s_0 es el nivel de recursos con el que comienza su vida (herencia recibida) y s_T el nivel con que la termina (herencias dadas). Por simplicidad se asume que la tasa de interés de la economía r es constante en el tiempo. En esta ecuación, el consumo intertemporal de una persona (el lado derecho) es igual a su ingreso permanente (el lado izquierdo). Nótese que los ingresos de capital no se incluyen en esta restricción intertemporal: cualquier ingreso de capital en un momento del tiempo refleja distintas elecciones acerca de cuándo consumir, y no diferencias en nivel de vida.

Mientras que en cada momento del tiempo el consumo y el ingreso pueden diferir, la ecuación anterior sugiere que es equivalente computar el nivel de vida intertemporal a partir de las fuentes de ingreso o de sus usos. En la práctica, ninguna de estas dos estimaciones resulta sencilla. La razón es obvia: para estimar el ingreso o consumo intertemporal hay que conocer estas variables en cada momento del tiempo, inclusive en períodos en el futuro. Los economistas no se han dejado intimidar por esta tarea ambiciosa y a menudo producen estimaciones tanto del lado izquierdo, como del derecho de la ecuación intertemporal. Naturalmente, para ello es necesario hacer supuestos.

La estimación del lado de las fuentes (lado izquierdo) requiere estimar los ingresos laborales a lo largo de la vida de la persona. Conociendo ciertas características de esa persona (ej. su nivel educativo, su género, su región de residencia) pueden predecirse sus ingresos futuros a partir de modelos econométricos. El apéndice IV del libro presenta una variedad de técnicas cuantitativas para estimar modelos de ingresos

laborales. Es preciso reconocer, sin embargo, que con las variables disponibles en una encuesta de hogares el poder predictivo de estos modelos es modesto y por consiguiente las estimaciones del ingreso permanente por el lado de las fuentes son en general débiles.

La estimación del lado de los usos (lado derecho) requiere conocer el consumo en cada período de tiempo, incluidas las herencias o regalos dejados. Típicamente quienes siguen este camino invocan acá un supuesto crucial: la gente suaviza su consumo a lo largo de la vida. En el extremo, si las personas mantienen un consumo estable en el tiempo, el nivel de consumo corriente de cualquier período se convierte en un buen indicador del ingreso o consumo intertemporal. El supuesto clave es la ausencia de restricciones de liquidez que impidan a una persona endeudarse o ahorrar libremente a una tasa de interés dada. En la realidad las restricciones de liquidez existen para períodos largos de tiempo, como puede comprobar casi cualquier joven latinoamericano que quiera endeudarse para comprar una casa y pagarla a 30 años. Las restricciones, en cambio, son menos rígidas en períodos más cortos. Es posible que el mismo joven consiga financiamiento (muy probablemente informal o familiar), si decide comprarse una bicicleta para ir a la escuela o al trabajo.

El análisis distributivo intertemporal no es usual en la práctica, ni en América Latina ni en el resto del mundo, por dos razones. Desde el lado conceptual se sostiene que socialmente interesan las carencias y padecimientos actuales, independientemente de que puedan ser superadas en alguna otra etapa de la vida. El argumento práctico se afirma en que la medición del nivel de vida corriente es más fácil y genera menos controversias metodológicas que la estimación del nivel de vida intertemporal. Por estas razones la literatura distributiva, y el resto de este libro, se concentran en la distribución de variables que aproximan el nivel de vida actual y no intertemporal.

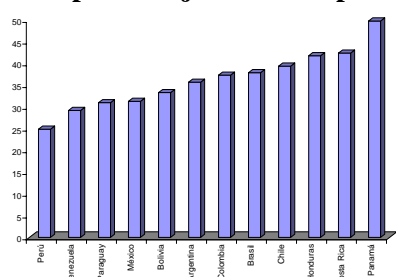
5. Personas, factores y grupos

La unidad de estudio en un típico análisis distributivo es la persona. En términos normativos nos interesa evaluar cuál es el nivel de desigualdad en el nivel de vida de las personas, y nos preocupan los individuos que no logran superar un determinado umbral de bienestar, independientemente de dónde viven, a qué grupo pertenecen o de cuáles factores productivos son propietarios. A las distribuciones de variables definidas al nivel de individuos se las conoce como distribuciones *personales*. En particular, gran parte de la literatura distributiva se ha concentrado en evaluar diferencias entre personas caracterizadas por su ingreso: entre pobres y no pobres; entre el 10% de mayores ingresos y el 10% de menores ingresos, etc.

En la distribución *funcional* el interés se traslada a los distintos factores productivos generadores de ingreso: trabajo, capital, tierra y otros. La unidad de análisis no es la persona, sino los factores de producción. Interesa, por ejemplo, saber cuál es la participación del capital o el trabajo en el ingreso nacional y cómo ha evolucionado en el tiempo.

La principal fuente de información para estimar las distribuciones personales del ingreso son las encuestas de hogares. Desafortunadamente, estas encuestas tienen serios problemas en captar adecuadamente los ingresos de capital, utilidades y rentas. Por esta razón, las estimaciones de las distribuciones funcionales están casi exclusivamente basadas en el sistema de Cuentas Nacionales. La figura 5.1 muestra estimaciones del indicador más usual de distribución funcional - la participación de la masa salarial en el ingreso o producto nacional – para varios países de América Latina, sobre la base de información de Cuentas Nacionales.

Figura 5.1
Participación de la remuneración de los asalariados
como porcentaje del PIB a precios corrientes de mercado



Fuente: CEPAL (BADEINSO). Datos correspondientes a 2000 o año cercano.

La figura muestra la participación del trabajo asalariado y no de todo el factor trabajo, dado que es difícil para el sistema de Cuentas Nacionales identificar el pago al trabajo por cuentapropia. Esta deficiencia resulta muy relevante, dada la importancia del auto-empleo en los países de la región. Por ejemplo, información de encuestas de hogares revela que en Bolivia sólo el 36% de los trabajadores eran asalariados en 2007.

Mientras que el grueso de la disciplina se ha concentrado en el análisis de la distribución personal, algunos analistas reclaman un papel más central para la distribución funcional, argumentando que las categorías de trabajadores y capitalistas están lo suficientemente polarizadas y la dispersión al interior de cada grupo es lo suficientemente poco relevante como para que sea la desigualdad entre trabajo y capital el centro natural de todo estudio distributivo. Si la inequidad social es ante todo desigualdad entre grupos diferenciados por su acceso a los medios de producción, es claro que la distribución funcional debe adquirir un papel central. En este enfoque la relación entre la distribución funcional y personal es bastante directa: un incremento de la participación de los salarios en el ingreso nacional es interpretado como una reducción de la desigualdad en la distribución personal del ingreso. Por otra parte, es común la asociación entre clase trabajadora y pobreza.

De hecho, en sus inicios la Economía se focalizó en el estudio de la distribución funcional, asociando la propiedad de factores productivos con clases sociales relativamente homogéneas internamente y claramente diferenciadas entre ellas en

términos de bienestar.³⁵ Con el tiempo, el interés se ha ido volcando hacia la distribución personal por diversas razones (Milanovic, 2011). Por un lado, en el siglo XX florecieron las encuestas de hogares, casi inexistentes antes, que permiten explorar con mucho mayor detalle la distribución personal. Por otro lado, las sociedades se complejizaron: la propiedad del capital y la tierra se ha ido desconcentrando (respecto de siglos anteriores) difundiéndose hacia estratos medios de la población; el capital humano adquirió un papel central, volviendo a los ingresos del trabajo una fuente fundamental de recursos y ascenso social; aumentó la dispersión salarial entre trabajadores hasta convertirse en un factor importante de la desigualdad de ingresos, y se expandieron los estratos medios en casi todas las sociedades. Estos fenómenos hacen que la visión de una sociedad polarizada entre asalariados (pobres) y capitalistas (ricos) sea una simplificación exagerada. En la actualidad la relación entre distribución funcional y personal se ha vuelto más difusa: buena parte de los individuos en la cima de la distribución personal son asalariados (ej. gerentes de empresas), mientras que hay personas de recursos medios y bajos que derivan una parte importante de sus ingresos de fuentes distintas al trabajo (uso del capital propio en auto-empleo, rentas).³⁶ Por supuesto, continúa siendo cierto que el capital y la tierra están más concentrados en los estratos superiores de ingresos de la población, por lo que la distribución funcional puede servir de aproximación a la distribución entre individuos. Sin embargo, existiendo los datos para medir directamente la distribución personal, no parece razonable seguir un camino indirecto de estimación.

Concentrarnos en la distribución funcional dificulta la medición y el análisis de la pobreza, y el estudio de muchos aspectos relevantes de la desigualdad, derivados de las diferencias de ingresos intragrupal y de la imperfecta correlación entre ingresos y propiedad de factores. Sin embargo, nótese que esta consideración no implica desechar el estudio de la distribución funcional, ya que su análisis puede ser útil para entender aspectos relevantes de la distribución personal.³⁷

Una segunda línea de defensa del estudio de la distribución funcional está basada en las deficiencias de las encuestas de hogares en captar ingresos de capital. Dada esta debilidad, las mediciones funcionales basadas en Cuentas Nacionales iluminan un aspecto de la desigualdad que las mediciones personales subestiman. Este argumento práctico es razonable, aunque no alienta a desestimar el estudio de la distribución personal, sino a mejorar o complementar sus estimaciones con la ayuda de información de Cuentas Nacionales. El Apéndice III trata este punto con más extensión.

³⁵ Ricardo (1817) en sus Principios señala que “La determinación de las leyes que rigen la distribución (funcional) es el problema primordial de la Economía Política”.

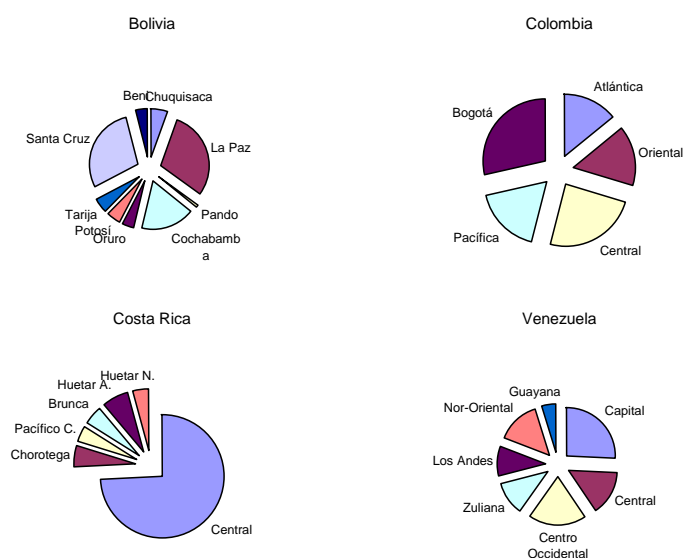
³⁶ Piketty y Saez (2006) reportan que en Estados Unidos los salarios ya se han convertido en la principal fuente de ingresos para el 1% más rico de la población. Atkinson, Piketty y Saez (2009) reportan evidencia sobre la creciente relevancia de los ingresos salariales en otros países.

³⁷ En este sentido, la literatura actual parece haber sobre-reaccionado en su énfasis por la distribución personal, otorgándole poco espacio al estudio y medición de la distribución funcional.

En síntesis, la distribución funcional es relevante y se justifica su medición y estudio. Sin embargo, compartimos el criterio de la literatura actual en otorgarle una categoría inferior a la que ocupa la distribución personal del ingreso. El estudio de esta última nos permite una caracterización conceptual y empírica más precisa de la pobreza y la desigualdad. El análisis de la distribución funcional puede ser un complemento importante para mejorar las estimaciones de la distribución personal, y para contribuir a entender su forma y cambios en el tiempo.

Así como en la distribución funcional se agrupa a las personas según sus factores de producción, es posible clasificarlas de acuerdo a cualquier otra variable, ignorando las diferencias internas y tratando a cada grupo como una unidad. Supongamos por ejemplo que se agrupa a las personas de acuerdo a su región de residencia. En ese caso la distribución se transforma en una lista de ingresos promedio de cada área geográfica: ésta es la *distribución regional* del ingreso. La figura 5.2 presenta información sobre la distribución regional en algunos países de América Latina, utilizando información de población e ingresos de las encuestas de hogares. Para algunos fines (ej. estudios de federalismo fiscal) la distribución regional es relevante *per se*; en otros casos puede contribuir al análisis de la distribución personal.

Figura 5.2
Distribución regional del ingreso



Fuente: elaboración propia en base a microdatos de encuestas de hogares.

6. Fuentes de información

En esta sección discutimos la procedencia de los datos para llevar a cabo un análisis distributivo. Idealmente sería óptimo contar con información de toda la población: un *censo* es la mejor fuente de datos para toda investigación empírica. Sin embargo, en la práctica los censos no son útiles para un trabajo distributivo debido a dos razones: se

realizan con poca frecuencia y suelen no incluir preguntas ni de ingreso ni de consumo. Los cuestionarios necesarios para relevar estas variables monetarias son largos y sobre temas que algunas personas se sienten incómodas en contestar, lo cual entra en conflicto con el objetivo central de un censo: recabar información demográfica, minimizando la no respuesta.³⁸

Las *encuestas* constituyen la principal fuente de evidencia empírica sobre temas sociales. En particular, el análisis distributivo se nutre de *encuestas de hogares*, es decir cuestionarios destinados a una muestra representativa de la población, entrevistada en sus viviendas, en la que se recaba información sobre todos los miembros del hogar, incluyendo datos de ingreso y/o consumo. Todos los países de América Latina cuentan con institutos oficiales de estadística que implementan periódicamente encuestas de hogares de las cuales se extraen microdatos para el estudio de la pobreza, la desigualdad y otros temas sociales.

Antes de profundizar sobre las encuestas, mencionemos algunas *fuentes indirectas* de información, cuyo objetivo explícito no es asistir a las estadísticas sociales. Un ejemplo es el de las declaraciones de impuestos a los ingresos. En todos los países existen gravámenes al ingreso (renta, ganancias) que exigen al contribuyente la declaración de los montos percibidos en el año, lo cual constituye información útil para un estudio distributivo. Naturalmente, esta fuente es relevante sólo en países donde la cobertura del impuesto es amplia. La importancia del impuesto a los ingresos en los sistemas tributarios latinoamericanos es escasa. La gran mayoría de la población de la región no completa una declaración jurada de impuestos, ya sea porque no le corresponde legalmente o porque desarrolla sus actividades en el sector informal de la economía. Por estas razones las declaraciones de impuestos, u otras fuentes indirectas como registros administrativos de salarios o declaraciones al sistema de seguridad social, no pueden servir de base informativa de un estudio distributivo comprehensivo en los países de América Latina, aunque a menudo pueden ser usados como fuentes complementarias. Por ejemplo, las declaraciones de impuestos pueden servirnos para conocer mejor los ingresos de las personas más ricas, las cuales usualmente no son incluidas en las encuestas de hogares (ver Apéndice III).³⁹ Las declaraciones de impuestos tienen la ventaja adicional de ser realizadas todos los años y en algunos países la información está disponible desde hace mucho tiempo, en contraste con los datos de otras fuentes más modernas, como las encuestas.

Volvamos a las encuestas de hogares, base de la mayoría de los estudios distributivos empíricos. Las siguientes secciones discuten los problemas de muestreo y representatividad. Proveer una discusión pormenorizada de estos temas excede los

³⁸ Algunos censos incluyen un módulo especial donde se incorporan preguntas de ingreso o consumo, pero se lo distribuye sólo a una pequeña fracción de la población, lo cual naturalmente hace que la información obtenida deje de ser censal. El Censo 2010 en Panamá incluyó un cuestionario de ingreso para toda la población, lo cual atentó contra la calidad del proceso.

³⁹ Ver Atkinson y Piketty (2006), Atkinson, Piketty y Saez (2009) para estudios en países desarrollados, y Alvaredo (2010) para una aplicación a un país de América Latina (Argentina).

objetivos de este libro. En su lugar, se presentan algunos conceptos básicos y referencias.⁴⁰

6.1. Muestreo

Una encuesta es una muestra representativa de la población. Sobre la base de un *marco muestral* que lista los integrantes de una población – típicamente el último censo nacional disponible en el país - el instituto de estadística nacional selecciona un procedimiento de muestreo y realiza el trabajo de campo encuestando a las personas escogidas de acuerdo al mecanismo de selección implementado.

El procedimiento más sencillo es el *muestreo aleatorio simple*, en el cual cada elemento de la población tiene la misma probabilidad de aparecer en la muestra. El simple azar determina quién es encuestado y quien no en la población, teniendo todos igual probabilidad de ser seleccionados. Desde un punto de vista formal no hay razones para adoptar mecanismos muestrales que no sean del tipo aleatorio simple. Sin embargo, por algunas razones prácticas estos esquemas son de uso infrecuente.

Estratificación

Supongamos que las autoridades requieren información no sólo del ingreso promedio nacional, sino también de los ingresos medios de cada provincia en la que está dividido el territorio, con el fin de evaluar diferencias regionales. Un esquema de muestreo aleatorio simple puede implicar que muy pocas observaciones provengan de alguna provincia poco poblada, con la consiguiente dificultad para obtener estimaciones precisas del ingreso medio de esa jurisdicción. Ante esta perspectiva resulta conveniente separar el marco muestral por provincias y tomar muestras de tamaño apropiado en cada una de ellas. Este es un caso de *muestreo estratificado*: la población es primero partida en conjuntos disjuntos o estratos (ej. provincias) y luego se toman muestras de tamaño suficiente dentro de cada estrato.

Supongamos que para garantizar estimaciones confiables del ingreso de una provincia habitada por sólo el 1% de la población nacional la muestra de la encuesta se estratifica, forzando a que el 5% de las observaciones provengan de esa provincia. Esta sobrerrepresentación es útil para mejorar las estimaciones provinciales, pero hace necesario algún ajuste a la hora de estimar el ingreso medio de la población nacional. La dificultad técnica consiste en “desandar” el proceso de estratificación a fines de

⁴⁰ La literatura estadística sobre el tema de muestreo es abundante, ver por ejemplo Levy y Lemeshow (1999) para un tratamiento básico y Sarndal *et al* (2003) para un análisis más sofisticado y preciso. Es curioso que mientras que el tema de muestreo ocupa un espacio significativo en la sociología, el marketing y, naturalmente, la estadística, su lugar en la currícula de los economistas es menor, máxime cuando una parte considerable del análisis empírico en economía aplicada se basa en datos obtenidos a través de complejos mecanismos muestrales. De hecho, la gran mayoría de los textos de econometría ignoran esta temática, siendo los libros de Deaton (1997) y Cameron y Trivedi (2003) algunas de las pocas excepciones.

restaurar la estructura probabilística de un proceso aleatorio simple. En nuestro ejemplo de muestreo estratificado los habitantes de la provincia pequeña aparecen en la muestra con mayor frecuencia que en la población (5% vs. 1%), de modo que es necesario otorgarles menor ponderación en el agregado, con el objeto de restituir su participación relativa en la población. La sección 6.2 elabora sobre este punto.

Clusters

Un *cluster* o conglomerado es un agrupamiento natural de la población. Por ejemplo, la gente no vive esparcida uniformemente en el espacio, sino que en su mayoría se agrupa en pueblos y ciudades, que constituyen agrupamientos o *clusters* poblacionales. La aplicación de un mecanismo completamente azaroso donde cada persona en la población tiene igual probabilidad de ser elegido (muestreo aleatorio simple) implicaría encuestar a personas dispersas en muchos pueblos y ciudades diferentes. Una alternativa consiste en realizar un muestreo en dos etapas: primero seleccionar al azar sólo algunos *clusters* poblacionales, y en la segunda etapa elegir aleatoriamente las personas a encuestar dentro de cada *cluster* previamente escogido.⁴¹ Este procedimiento implica un ahorro de costos de implementación, al tener que visitar sólo algunos pueblos y ciudades. Por ejemplo, resulta más económico obtener una muestra de 100 hogares visitando 10 pueblos y encuestando a 10 hogares en cada uno, que visitando 100 pueblos y encuestando a un hogar en cada uno. Este ahorro se obtiene a costa de una clara desventaja: las observaciones dentro de un *cluster* no son independientes. Se espera, por ejemplo, que las características de los vecinos de un pueblo, entre ellas el ingreso, estén correlacionadas.

El agrupamiento introduce dependencias que a veces son de difícil tratamiento empírico, y que en general inducen a errores estadísticos significativos. Intuitivamente, la dependencia entre las observaciones resta variabilidad a la muestra, lo que lleva a subestimar la verdadera varianza si esta dependencia es omitida. Dicho de otra forma, 10 observaciones independientes son en general más informativas que 10 observaciones altamente dependientes. En este sentido en la práctica el proceso de *clusterización* tiende a aumentar la varianza de las estimaciones, en relación a un mecanismo de muestreo al azar. El muestreo por *clusters* funciona bien cuando las diferencias entre grupos son menores (de modo que se pierde poca información por ignorar los *clusters* no seleccionados) y las diferencias dentro del *cluster* son grandes.

Nótese una diferencia importante entre estratos y *clusters*: los primeros son definidos en forma determinística y su unión agota la población, mientras que los *clusters* son escogidos aleatoriamente y constituyen sólo un subconjunto de la población.

La estratificación y el muestreo por *clusters* no son alternativas sustitutas, sino complementarias, a menudo aplicadas conjuntamente. En la práctica las encuestas de

⁴¹ Este proceso implica la existencia de unidades primarias de muestreo (los clusters elegidos) y unidades secundarias de muestreo (los hogares seleccionados dentro de los clusters).

hogares aplican *diseños muestrales complejos* que involucran varias etapas de estratificación y clusterización. Por ejemplo, un esquema muestral podría iniciarse dividiendo toda la población en estratos definidos en función de regiones geográficas, eligiendo luego en cada estrato una muestra de *clusters*, y finalmente practicando un muestreo aleatorio simple dentro de cada *cluster*.

Un ejemplo de esquema muestral: la CASEN de Chile ⁴²

El marco muestral de la CASEN es el Censo de Población y Vivienda de Chile. El país se divide en regiones, provincias y comunas, las cuales son a su vez subdivididas a fines censales. El muestreo usado en la CASEN es estratificado, por *clusters* y de múltiples etapas. La estratificación es geográfica, dividiendo el país en 553 estratos, en base a la combinación de divisiones político-administrativas (comunas o agrupaciones de comunas) y áreas (urbana o rural). Hay dos tipos de comunas: auto-representadas y co-representadas. En las primeras el número de observaciones en la muestra es suficiente para realizar estimaciones independientes. El tamaño muestral de las co-representadas, en cambio, sólo permite estimaciones a nivel agregado.

El diseño muestral es de dos etapas para las comunas auto-representadas. Las unidades primarias de muestreo (UPM) son los sectores de empadronamiento censal (conglomerados de viviendas). En el área urbana se refieren a una manzana o parte de ella, mientras que en el área rural están definidos como un conjunto de viviendas próximas. En ambos casos contienen un número de viviendas tal que el empadronador del censo pueda encuestarlas en un día. Las UPM se seleccionan con probabilidad proporcional al tamaño (número de viviendas particulares obtenidas a partir del Censo). Las unidades secundarias de muestreo son las viviendas particulares ocupadas en forma permanente. Las viviendas son seleccionadas de manera aleatoria.

La gran mayoría de los estudios distributivos ignoran los problemas de muestreo, asumiendo implícitamente un procedimiento aleatorio simple, lo cual genera sesgos en las estimaciones de las varianzas de los estadísticos utilizados. Tres razones han justificado este proceder: (i) a menudo es difícil obtener información técnica de los institutos de estadística sobre los procedimientos de muestreo, y las variables incluidas en las bases de datos son insuficientes para realizar ajustes en las estimaciones;⁴³ (ii) el ajuste de las varianzas al incorporar muestreos complejos es técnicamente engorroso;⁴⁴

⁴² La CASEN es la Encuesta de Caracterización Socioeconómica Nacional de Chile. Este recuadro está basado en Tornarolli (2008), donde se describen con mayor detalle los esquemas muestrales de varias encuestas de América Latina.

⁴³ Adicionalmente, cuando existe, la información tiene diferente grado de detalle entre países, lo que implica que la inclusión de ajustes no necesariamente mejora la comparabilidad.

⁴⁴ Estas técnicas incluyen linealizaciones de Taylor y/o procedimientos basados en réplicas. Ver Sarndal *et al.* (1992) para un tratamiento pormenorizado, si bien técnicamente más demandante.

y (iii) con frecuencia las estimaciones que consideran estos ajustes no difieren sustancialmente de aquellas que los ignoran.

De a poco estas razones están perdiendo validez, ya que es creciente el número de oficinas de estadística que brindan información sobre procedimientos muestrales y existen cada vez más instrumentos computacionales que permiten acomodar los requisitos de estructuras muestrales complejas. Por ejemplo, las últimas versiones de Stata han incorporado una gran variedad de herramientas para el análisis de datos provenientes de muestreos complejos.⁴⁵ La próxima sección incluye un ejemplo sobre la importancia de tomar en cuenta el tipo de muestreo del que provienen los datos de las encuestas de hogares.

Tamaño de muestra y representatividad

Una pregunta inevitable (e incómoda) que surge en todo trabajo empírico es ¿cuál es el tamaño de la muestra necesario para estimar un indicador? Consideremos un ejemplo sencillo. Supongamos que el interés consiste en estimar la tasa de pobreza, es decir el porcentaje de personas pobres. A fines de simplificar el análisis, supondremos que disponemos de una muestra proveniente de un esquema aleatorio simple.⁴⁶ ¿Cuál es el tamaño de muestra que hace falta para estimar la tasa de pobreza de la población? Interpretada literalmente, la respuesta a esta pregunta es trivialmente uno. Con una sola persona es posible estimar la tasa de pobreza, aunque claramente las estimaciones serán muy poco confiables. Cuanto más grande es el tamaño de la muestra mayor será la confiabilidad de la estimación. En consecuencia, la respuesta a la pregunta anterior dependerá del grado de confiabilidad que se requiera para la estimación. Llevemos el caso a un extremo. Supongamos que pretendemos máxima confiabilidad, es decir, no toleraríamos ninguna discrepancia entre nuestra estimación y la verdadera tasa de pobreza. En este caso la respuesta acerca del tamaño óptimo es también trivial: el tamaño de la muestra deberá coincidir con el de la población. Es la única garantía de que no haya discrepancias entre la estimación y el verdadero parámetro poblacional.

En la práctica, en los casos no extremos el problema del tamaño óptimo de la muestra tiene solución no-trivial. Consideremos un ejemplo simple adaptado de Levy y Lemeshow (1999). Supongamos que estamos interesados en estimar la tasa de pobreza H de una comunidad, en un esquema de muestreo aleatorio simple. El estimador a usar es la proporción de personas que reportan ingresos inferiores a un umbral o línea de pobreza. Supongamos que las personas que encargan la encuesta dicen no querer tolerar más de un 7% de discrepancias entre la estimación y el verdadero valor de la tasa de

⁴⁵ Kreuter y Valliant (2009) es una buena referencia para el uso de comandos de muestreo en Stata. SUDAAN es una herramienta específica para las cuestiones muestrales.

⁴⁶ Asumimos también que hay heterogeneidad entre las personas. Trivialmente, si sabemos que en una comunidad todas las personas son iguales, y por ende o bien son todas pobres o ninguna lo es, basta con encuestar a una sola persona para inferir el status de pobreza de toda la población. Con variabilidad muestral nula, el tamaño necesario de la muestra es 1.

pobreza.⁴⁷ A fines de determinar un tamaño de muestra adecuado para satisfacer este requisito, recordemos que un intervalo de confianza al 95% por ciento contiene a la verdadera proporción de personas pobres con probabilidad 95%. Es decir, es “casi cierto” que el intervalo $\hat{H} \pm 1.96 \cdot \text{std}(\hat{H})$ contiene a la verdadera tasa de pobreza. En este caso, \hat{H} es la estimación de la tasa de pobreza H , 1.96 es el percentil 0.975 de la distribución normal estándar y $\text{std}(\hat{H})$ el error estándar de \hat{H} . Nótese que requerir que no haya más que un 7% de discrepancias (positivas o negativas) entre la estimación y el verdadero valor de la tasa de pobreza implica que el ancho del intervalo ($2 \cdot 1.96 \cdot \text{std}(\hat{H})$) no debe exceder a la verdadera tasa de pobreza en más de 7%. Entonces, nuestro problema es encontrar el tamaño de muestra que satisface $2 \cdot (1.96 \cdot \text{std}(\hat{H})) \leq 2 \cdot (0.07H)$. Utilizando la fórmula para la varianza asintótica de \hat{H} , recordemos que $\hat{V}(\hat{H}) = \hat{H}(1 - \hat{H})/n$.

Despejando n en la fórmula anterior obtenemos

$$(6.1) \quad n \geq \left[\frac{1.96 \sqrt{\hat{H}(1 - \hat{H})}}{H \cdot 0.07} \right]^2$$

Nótese que la verdadera tasa de pobreza H no es observada. En la práctica, el tamaño óptimo de la muestra para estimarla se determina en base a alguna presunción o estimación previa. Si, por ejemplo, creemos que la tasa de pobreza ronda el valor 0.2, el tamaño de muestra requerido será mayor que 3136 personas. Si se aumentase la precisión de 0.07 a 0.05, entonces el tamaño requerido aumenta sustancialmente a 6147 personas.

En este ejemplo sencillo el “tamaño de muestra necesario” surge de especificar de antemano cuál es un nivel de “confiabilidad” (nuestra noción de “casi certeza”, en nuestro caso, 0.95), cuál es el nivel de precisión deseada (0.07) y nuestra presunción de la verdadera tasa de pobreza (0.2). La fórmula usada vale para un tipo de estimador (la proporción muestral); la solución se complica si se utilizan estimadores o esquemas de muestreo más complejos. Referiremos a Levy y Lemeshow (1999) para una discusión más pormenorizada.

6.2. Ponderadores

Dado que una encuesta es una muestra representativa de la población, las respuestas de cada individuo entrevistado representan las respuestas de un conjunto de personas en la población. Cada observación i en la encuesta tiene asociado entonces un *ponderador* o *factor de expansión* ω_i que indica el número de elementos en la población que la observación representa. El ponderador de una observación es igual (o proporcional) a la inversa de la probabilidad de ser muestreado. Por ejemplo, si en una población de un

⁴⁷ Por ejemplo, si H fuera 0.20, tolerar discrepancias del 7% implica exigir que el estimador muestral caiga en el intervalo [0.186, 0.214].

millón de habitantes se entrevista por muestreo aleatorio simple a 5000 personas, la probabilidad de ser encuestado sería 0.005 y el factor de expansión 200.

La tabla 6.1 divide a los hogares encuestados en cada país en quintiles de acuerdo al ingreso per cápita familiar, y reporta el valor promedio del factor de expansión para cada quintil. Dos puntos interesantes para observar: (i) los ponderadores difieren entre países; por ejemplo son superiores en un país muy poblado con una encuesta relativamente pequeña como México, y (ii) en cada país los ponderadores son en general crecientes en el nivel de ingreso (aunque no siempre uniformemente crecientes).

Cuadro 6.1

Ponderadores

Valores promedio por quintiles de ingreso

quintil	Argentina	Costa Rica	México	Nicaragua
1	213	81	1,021	114
2	229	89	1,274	126
3	227	94	1,348	136
4	237	99	1,341	156
5	237	116	1,320	168

Fuente: elaboración propia en base a microdatos de encuestas de hogares, circa 2008.

Hay varias razones prácticas que implican ponderadores heterogéneos. La primera es que los países implementan muestreos complejos que implican variabilidad en la probabilidad de muestreo. Por ejemplo, costos más elevados de recolección de información pueden reducir el muestreo en ciertas áreas geográficas, aumentando el factor de expansión de las unidades seleccionadas. La encuesta de hogares puede tener objetivos secundarios que exigen un muestreo más intenso en ciertas regiones o grupos. Por ejemplo, si el gobierno decide aprovechar la encuesta para estudiar con más detalle el impacto de ciertos programas sociales concentrados en un área geográfica, puede decidir incrementar el número de observaciones en esa región, para obtener mediciones más precisas. Consecuentemente, el factor de expansión de cada unidad observada en esa región será inferior al del resto del país. Una razón adicional de la no uniformidad de los ponderadores es discutida en el Apéndice III: la no respuesta a la encuesta. Si hay personas que se niegan a contestar el cuestionario y el instituto de estadística no está a tiempo de reemplazar esas observaciones por otras, debe incrementar la ponderación de los individuos “cercaños” que sí contestaron la encuesta, para que así el número total de observaciones expandidas alcance el valor de la población. Supóngase dos regiones con la misma población y con el mismo procedimiento de muestreo. Si en una región la tasa de no respuesta a la encuesta es alta, el valor del ponderador de cada observación será alto, en relación a aquellas observaciones en la región donde la tasa de no respuesta es menor.

Para evaluar la relevancia de incluir los ponderadores en los cálculos distributivos tomemos un indicador simple: la tasa de incidencia de la pobreza H , o proporción de

personas pobres. Teniendo en cuenta los ponderadores ω_i asociados a cada observación, esta tasa puede calcularse como

$$(6.2) \quad H = \frac{\sum_{i=1}^n 1(x_i < z) \omega_i}{\sum_{i=1}^n \omega_i}$$

donde i indexa a las n personas incluidas en la encuesta y $1(x_i < z)$ vale 1 si el ingreso x_i es inferior a la línea de la pobreza z y 0 en otro caso. En la ecuación cada término es multiplicado por el ponderador ω_i . Si los factores de expansión coinciden, es decir $\omega_i = \omega$, es sencillo notar que desaparecen del cálculo de (6.2): la tasa de pobreza resulta idéntica ya sea el cómputo se realice con o sin ponderadores. Si, en cambio, ω_i difiere entre individuos, los resultados ya dejan de ser idénticos. Supongamos un ejemplo muy sencillo de una sociedad en la que se encuesta a dos personas: la primera, que representa a 80 individuos, resulta pobre, y la segunda, que representa a 100 personas, tiene ingresos superiores a la línea de la pobreza. La tasa de pobreza H es 0.5 si se ignoran los ponderadores, mientras que es 0.44 si se los tiene en cuenta.

El cuadro 6.2 muestra las tasas de pobreza en las regiones de México, realizando el cálculo con y sin factores de expansión. Las tasas de pobreza suelen ser menores cuando se incluyen los ponderadores, dado que los factores de expansión son en general superiores en los estratos de mayores ingresos, donde las tasas de no respuesta son más elevadas.

Cuadro 6.2
Tasa de pobreza en México, 2006
Cálculos alternativos usando o no los ponderadores

Región	Uso de ponderadores	
	No	Si
Noroeste	10.0	9.7
Norte	18.7	13.4
Noreste	11.4	6.4
Centro-Occidente	12.5	13.2
Centro-Este	14.8	10.7
Sur	45.6	35.9
Oriente	22.4	21.9
Península de Yucatan	19.8	15.7
Nacional	18.8	14.8

Fuente: elaboración propia en base a microdatos de la ENIGH, 2006.

La recomendación práctica es usar siempre que sea posible los factores de expansión, obteniendo así estadísticas ponderadas que constituyen una mejor aproximación a la realidad de la población total. A fines descriptivos y cuando el objetivo es realizar

conjeturas acerca de una colección de objetos de referencia (la población), el uso de ponderadores es relevante.⁴⁸

Analicemos un ejemplo concreto para ilustrar el impacto de considerar el diseño muestral sobre los resultados. Para ello empleamos la Encuesta de Hogares de Propósitos Múltiples (EHPM) de Costa Rica en 2006, donde la muestra se selecciona en dos etapas a partir de una estratificación por regiones geográficas (6 regiones) y zonas (urbanas y rurales).⁴⁹ En la primera etapa del muestreo se seleccionan las Unidades Primarias de Muestreo (UPM) a partir de segmentos censales. Finalmente, las Unidades Secundarias de Muestreo son viviendas seleccionadas dentro de cada UPM. Como ejemplo, se estima para cada quintil de ingreso laboral la proporción de trabajadores empleados en industrias básicas (ver cuadro 6.3).⁵⁰ El panel (a) ignora el diseño muestral, en el (b) sólo se consideran los factores de expansión o ponderadores, el (c) agrega la consideración de la estratificación de la muestra, mientras que el panel (d) incorpora también la estructura de *clusters*.

Como puede observarse la utilización de ponderadores impacta considerablemente sobre la estimación puntual y los errores estándar de la proporción de individuos empleados en el sector de industrias básicas. En este caso particular, considerar la estructura de estratificación – además de los ponderadores -- casi no modifica los errores estándar. Por su parte, cuando se agrega la “clusterización” se obtienen errores estándar sustancialmente más elevados.

El efecto del diseño muestral sobre las estimaciones de la varianza puede medirse mediante el “efecto diseño”, que se calcula como el cociente entre la varianza estimada teniendo en cuenta el diseño muestral y la varianza que se obtendría si la muestra hubiera sido obtenida mediante un muestreo aleatorio simple. En el cuadro 6.3 dicho indicador aparece en las columnas indicadas con *deff*, que abrevia “*design effects*”.

Cuadro 6.3

Efecto sobre estimaciones puntuales y errores estándar de considerar el diseño muestral
Proporción de trabajadores en industrias básicas
Costa Rica, 2006

⁴⁸ La literatura econométrica en cambio es cauta en relación al uso de ponderadores. Cameron y Trivedi (2005) sugieren que el uso de ponderadores es necesario cuando se utiliza un enfoque descriptivo o de resumen de datos de una población concreta de la cual se obtuvo una muestra. En cambio, si el interés recae en estimar características de una relación modelística (ej. regresiones) no es necesario utilizar ponderaciones si la especificación del modelo es correcta. Dichos autores enfatizan que esta apreciación vale siempre y cuando la estratificación no sea endógena.

⁴⁹ Las seis regiones identificadas en la encuesta son Región Central, Chorotega, Pacífico Central, Brunca, Huetar Atlántica y Huetar Norte. Las características del diseño muestral de la Encuesta de Hogares de Propósitos Múltiples de julio 2006 se describe en INEC (2006).

⁵⁰ En esa categoría incluimos a las industrias de alimentos, bebidas y tabaco, y textiles y confecciones.

(a): sin ajustes				(b): ajuste por ponderadores			
quintil	media	desvío	deff	quintil	media	desvío	deff
1	0.0648	0.00414	1.00	1	0.0643	0.00465	1.15
2	0.0736	0.00459	1.00	2	0.0854	0.00584	1.40
3	0.0720	0.00451	1.00	3	0.0719	0.00500	1.20
4	0.0662	0.00438	1.00	4	0.0597	0.00444	1.13
5	0.0467	0.00401	1.00	5	0.0459	0.00454	1.51

(c): ajuste por ponderadores y estratos				(d): ponderadores, estratos y clusters			
quintil	media	desvío	deff	quintil	media	desvío	deff
1	0.0643	0.00465	1.15	1	0.0643	0.00540	1.56
2	0.0854	0.00584	1.40	2	0.0854	0.00682	1.91
3	0.0719	0.00499	1.20	3	0.0719	0.00646	2.00
4	0.0597	0.00443	1.12	4	0.0597	0.00544	1.69
5	0.0459	0.00454	1.51	5	0.0459	0.00602	2.65

Fuente: elaboración propia sobre la base de microdatos de la EHPM, Costa Rica 2006.

En el resto del libro, y salvo que se indique lo contrario, presentamos estimaciones que tienen en cuenta la estructura de ponderadores. En cambio, dado lo dificultoso que resulta obtener una descripción detallada del diseño muestral de todas las encuestas de América Latina, y para no complicar la exposición de la implementación práctica de los resultados, en general ignoramos el ajuste de las varianzas de los estimadores al considerar esquemas muestrales complejos.

6.3. Diseño y recolección

El diseño de la encuesta constituye una de las etapas cruciales en la obtención de evidencia empírica sobre un fenómeno. Las mediciones de pobreza y desigualdad dependen fuertemente de las preguntas incluidas (y las omitidas) en los cuestionarios de las encuestas. El diseño de una encuesta de hogares exige alcanzar un cuidadoso balance entre la ambición por cubrir un amplio rango de temas, y la prudencia en la longitud y complejidad del cuestionario. Por un lado una encuesta nacional es una oportunidad única para conocer la realidad social del país, y por ende es normal tratar de aprovecharla para obtener información sobre tópicos específicos. Por otro lado, cuestionarios largos con preguntas complejas aumentan sustancialmente la tasa de no respuesta, o el número de respuestas inconsistentes o no confiables, y por consiguiente comprometen el éxito de toda la encuesta.

En este libro tomamos la perspectiva del usuario de datos, en lugar del generador de información, y en consecuencia no abordamos los importantes temas de diseño de encuestas y recolección de datos. Con frecuencia, sin embargo, el analista de temas distributivos se ve en la situación de tener que diseñar un cuestionario y una estrategia de muestreo y recolección de información. En los últimos años hay una clara tendencia hacia el aumento de las encuestas *ad hoc* para estudiar temas sociales específicos.

Siendo la encuesta uno de los instrumentos centrales en la obtención de información sobre la realidad, y no pudiendo las encuestas de hogares oficiales nacionales cubrir todos los temas de interés, son cada vez más los sondeos sobre temas específicos implementados por organismos internacionales, centros de investigación, empresas u oficinas del gobierno. El conjunto de herramientas de un analista social debe incluir conocimientos básicos sobre diseño de encuestas.⁵¹

6.4. Tipos de encuestas de hogares

Existen dos clases principales de encuestas de hogares útiles para un análisis distributivo: las encuestas de empleo e ingresos, y las de condiciones de vida. En lo que sigue se presentan brevemente las principales características de cada tipo, ejemplificando con casos concretos en América Latina.

Las encuestas más extendidas en los países latinoamericanos son aquellas cuyo principal objetivo es relevar las condiciones de empleo e ingresos de una muestra representativa de la población. Casi todos los países de la región cuentan con un sistema periódico de encuestas de empleo e ingresos (EEI), cuya principal utilidad es servir de base para las estadísticas oficiales de desempleo, y en muchos casos pobreza. Estas encuestas suelen incluir también información valiosa sobre las características demográficas, habitacionales y educativas de la población, y un cuestionario destinado a relevar todas las fuentes de ingreso de todos los miembros del hogar. El acento, sin embargo, está puesto en la situación laboral de las personas, por lo que estas encuestas dedican mucho espacio a captar la condición de empleo, las características del puesto de trabajo, los ingresos laborales y el tipo de desempleo. En el Apéndice II del libro se detalla la estructura de una típica EEI ejemplificando con el caso de la *Pesquisa Nacional por Amostra de Domicílios*, la principal encuesta de hogares de Brasil.

Las encuestas de condiciones de vida (ECV) son más extensas y ambiciosas que las EEI, suelen incluir información sobre gastos de consumo y contienen secciones sobre un amplio conjunto de temas sociales. Dado que relevan datos de consumo, son usadas frecuentemente para construir canastas básicas de alimentos y líneas de pobreza. Las ECVs en general incluyen información de educación, salud, antropometría, empleo, migraciones, agricultura, ahorro y crédito, así como información recabada en cuestionarios comunitarios (no a hogares) sobre disponibilidad de servicios y precios locales.⁵² Dado que son significativamente más costosas, estas encuestas multipropósito se realizan con menos frecuencia. Mientras que las EEIs son típicamente anuales o continuas, las ECVs se realizan en forma más espaciada.

Las encuestas ECV fueron impulsadas por el programa *Living Standard Measurement Surveys* (LSMS) del Banco Mundial, establecido en 1980 con el objeto de mejorar la

⁵¹ Recomendamos Grosh y Glewwe (2000) como paso inicial para iniciarse en estos temas.

⁵² Grosh y Glewwe (2000) y Deaton y Grosh (2000) son dos referencias para conocer más acerca de estas encuestas.

disponibilidad de información social para países en desarrollo. Con el auspicio de ese programa se han realizado ECVs en varios países de América Latina (ej. Brasil, Ecuador, Guatemala, Nicaragua, Panamá, Perú). Otros países latinoamericanos han realizado ECVs y cuasi-ECVs con cuestionarios más cortos (usualmente sin información de gastos de consumo) fuera del programa LSMS. La CASEN de Chile y la ECV de Argentina corresponden a esta categoría. En el Apéndice II del libro se detalla la estructura de una ECV típica siguiendo el ejemplo de la Encuesta de Medición de Niveles de Vida (EMNV) 2005 de Nicaragua.

Además de las EEIs y las ECVs varios países realizan otras encuestas de carácter oficial. Son comunes las encuestas de gastos de los hogares (EGH), destinadas a relevar las canastas de consumo de la población, con la principal finalidad de construir los ponderadores para el índice de precios al consumidor (IPC), el principal índice usado para monitorear la inflación y brindar información al sistema de Cuentas Nacionales. A menudo, con los datos de estas encuestas de gastos se construyen canastas de consumo alimentario que sirven para calcular las líneas de pobreza oficiales de los países. Dado que relevan el consumo, y que con frecuencia incorporan secciones con preguntas sobre temas sociales, a veces las EGHs pueden confundirse con las ECVs. Las EGH tienen sin embargo una finalidad distinta y se realizan de forma más esporádica. Dado que el interés fundamental está en captar medias (ej. el consumo promedio de cada bien), en general tienen períodos de referencia más cortos, suficientes para la captación de medidas de tendencia central, pero inconvenientes para estudios distributivos por la variabilidad espuria que generan.

Algunos países realizan encuestas laborales frecuentes con el objeto de monitorear de manera ágil la dinámica del mercado laboral. Se trata de sondeos frecuentes (ej. mensuales) realizados a muestras de trabajadores o empresas con preguntas casi exclusivamente referidas a la situación en el mercado laboral. Estas encuestas laborales no deben confundirse con las EEIs, dado que en general no incluyen información de ingresos no laborales del hogar, y por ende no son adecuadas para monitorear la pobreza o la desigualdad de toda la población, aunque pueden utilizarse para estudiar desigualdad en el mercado laboral. La Pesquisa Mensal de Emprego de Brasil, la Encuesta Nacional de Empleo de Chile, la Encuesta Nacional de Ocupación y Empleo de México y la Encuesta Permanente de Empleo de Perú están entre las principales encuestas de este tipo en América Latina.

Casi todos los países implementan adicionalmente encuestas destinadas a estudiar alguna problemática social relevante. Nicaragua realizó una Encuesta de Discapacidad (ENDIS) en 2003; Paraguay una Encuesta de Hogares Indígenas en 2007/2008; Chile implementa cada tres años la Encuesta Nacional de Juventud; mientras que México lleva adelante anualmente la Encuesta de Evaluación de Hogares Urbanos (ENCELURB), para evaluar el programa Oportunidades, e implementa la Encuesta Nacional de Salud y Nutrición (ENSANUT). Estos son sólo algunos ejemplos de una larga y creciente lista de encuestas específicas.

Las encuestas de hogares nacionales no son las únicas fuentes de información distributiva. De forma creciente, en casi todos los países de América Latina se realizan encuestas de hogares destinadas a estudiar algún tema particular que no puede ser adecuadamente analizado con las encuestas oficiales existentes. Se trata de sondeos realizados por centros de investigación, organismos internacionales u ONGs usualmente con una muestra más pequeña que la de las encuestas nacionales. Si bien para monitorear pobreza y desigualdad estas encuestas son dominadas por las oficiales, pueden ser de utilidad para estudiar algunos temas específicos. Por ejemplo, en 2002 el Banco Mundial llevó a cabo la Encuesta de Impacto Social de la Crisis en Argentina (ISCA), de cobertura nacional con una muestra de 2800 hogares. La encuesta brinda valiosa información sobre las estrategias implementadas por los hogares para sobrevivir a la profunda crisis económica que experimentó ese país en 2001/02.⁵³

Las encuestas de opinión internacionales pueden ser utilizadas para algunos aspectos del análisis distributivo. Se trata en general de sondeos que indagan sobre opiniones y percepciones de la población, pero que suelen reportar información demográfica, educativa, laboral, de posesión de activos y en ocasiones de ingresos. Latinobarómetro es el principal ejemplo en la región. Se trata de un estudio de opinión pública que realiza desde 1995 alrededor de 19000 entrevistas anuales en 18 países de América Latina.⁵⁴ Gallup realiza una encuesta anual – Gallup World Poll – que cubre todos los países de América Latina y la gran mayoría de los países del mundo e incluye una pregunta de ingresos.⁵⁵ Si bien estas encuestas tienen muestras pequeñas en cada país (alrededor de 1000 observaciones), y aplican un procedimiento muestral no tan riguroso como los institutos nacionales de estadística, presentan algunas ventajas significativas: incluyen preguntas novedosas, hacen pública la información rápidamente y realizan las mismas preguntas en todos los países. Esta ventaja es muy relevante, si se recuerda que la principal fuente de problemas de comparabilidad internacional en las estadísticas sociales proviene de las diferencias en los cuestionarios entre países.

Las encuestas de hogares proveen datos de *corte transversal*, es decir información sobre un conjunto de personas en un momento dado del tiempo. Esta “foto” podría repetirse con las mismas personas año a año, de modo de obtener una “película” que nos permita estudiar problemas dinámicos, como los de movilidad o persistencia de la pobreza. Las encuestas con datos de *panel* están diseñadas para seguir a un conjunto de personas a lo largo del tiempo. Se trata, sin embargo, de encuestas caras. El encuestado se cansa de ser entrevistado año tras año, se muda, o se muere, con lo que el desgaste de la muestra es alto, lo cual requiere una muestra inicial más grande, y mayores esfuerzos en conservar las unidades de observación. El panel más famoso es el PSID (*Panel Study of Income Dynamics*) de Estados Unidos, llevado a cabo por la Universidad de Michigan

⁵³ Ver Fiszbein, Giovagnoli y Thurston (2003)

⁵⁴ www.latinobarometro.org

⁵⁵ Gasparini y Gluzmann (2011) indagan sobre la posibilidad de analizar pobreza y desigualdad en América Latina con la encuesta mundial de Gallup. Gasparini, Marchionni, Olivieri y Sosa Escudero (2011) estudian pobreza multidimensional con esa encuesta.

desde 1968. En los países de América Latina los paneles son o bien inexistentes, o bien de poca extensión. En los *paneles cortos* o *rotativos* se releva información repetida de un mismo hogar durante un período de tiempo poco prolongado; usualmente no más de dos años. La Encuesta Permanente de Hogares de Argentina, por ejemplo, entrevista a un hogar durante dos trimestres sucesivos, lo deja libre los dos trimestres siguientes, lo vuelve a entrevistar dos trimestres más, para luego eliminarlo definitivamente de la muestra. Las encuestas de hogares ENEMDU de Ecuador y ENAHO de Perú tienen una estructura de panel rotativo desde 2003. La mayoría de los países de la región ha implementado y prevé implementar encuestas de paneles cortos.⁵⁶

Además de algunos paneles cortos nacionales, existen en la región esfuerzos aislados de recolectar información de panel para proyectos específicos. Por ejemplo, la encuesta BASIS relevó información sobre un mismo conjunto de hogares rurales en El Salvador en los años 1995, 1997, 1999 y 2001, y sirvió para estudiar persistencia de ingreso y trampas de pobreza en ese país centroamericano.⁵⁷

6.5. Las encuestas de hogares en América Latina

Aunque las encuestas de hogares se remontan en Europa al siglo XVIII, no es hasta después de la Segunda Guerra Mundial en que los países comenzaron a organizar sistemas estadísticos que contemplan la realización periódica de estas encuestas. En la región, México fue el primero en adherirse a esta práctica en la década del 50. La mayoría de los países latinoamericanos o bien introdujo, o bien consolidó sus encuestas de hogares recién en la década del 70. Durante la década del 90 los países mejoraron sus encuestas de hogares, varios bajo el auspicio del programa MECOVI, un esfuerzo conjunto del Banco Mundial, CEPAL y el BID por mejorar las encuestas de la región. Los avances se manifestaron en una mayor frecuencia de encuestas, extensión de la cobertura hasta áreas rurales o regiones remotas, mejoramiento del cuestionario, inclusión de módulos especiales y mejoras en la captación del ingreso y/o el consumo.⁵⁸

Hoy en día todos los países de América Latina cuentan con oficinas estadísticas nacionales que implementan con periodicidad encuestas de hogares destinadas, entre otros motivos, a obtener estimaciones sobre pobreza y desigualdad. Todos los países realizan encuestas periódicas donde se releva el ingreso, y con menos frecuencia implementan cuestionarios de consumo. La mayoría de las naciones de la región han llevado a cabo encuestas de condiciones de vida, aunque en general de manera no sistemática.

⁵⁶ Existen proyectos más ambiciosos. Por ejemplo, en Chile se realizó un seguimiento a una muestra representativa de los hogares de las regiones III, VII, VIII y Metropolitana a partir de la encuesta CASEN 1996, tanto en 2001 como en 2006.

⁵⁷ Ver Rodríguez-Meza and González-Vega (2004) y Sosa Escudero, Marchionni y Arias (2006).

⁵⁸ El proceso de mejoras es obviamente elogiable, pero pone al analista en dificultades, dado que los cambios en cualquier característica de la encuesta afectan los resultados y dañan la comparabilidad en el tiempo, exigiendo mucho cuidado en el procesamiento de la información.

El cuadro 6.4 lista las principales encuestas usadas en este libro para medir pobreza y desigualdad en los 18 países de América Latina.⁵⁹ Estas son, a su vez, las encuestas utilizadas a lo largo del libro para ilustrar los distintos temas con evidencia empírica basada en microdatos. La mayoría de los países monitorea la pobreza y la desigualdad monetarias con encuestas de empleo e ingresos (EEI) con frecuencia anual, que incluyen información sólo de ingresos. En ese grupo se incluye a Argentina, Brasil, Costa Rica, República Dominicana, Honduras, Panamá, Paraguay, Uruguay y Venezuela.⁶⁰ Un conjunto de países incluye en este tipo de encuestas cuestionarios para relevar el consumo: Bolivia, Colombia, El Salvador, México y Perú entran dentro de esta categoría. En Nicaragua y Guatemala el monitoreo de la pobreza está basado en encuestas ECV con datos de ingreso y consumo, mientras que Ecuador combina el uso de varias ECVs con una EEI de cobertura nacional reciente (ENEMDU). Finalmente, Chile monitorea las variables distributivas con la CASEN, una cuasi-ECV sin información de consumo.

En 2011 todas las encuestas de América Latina tenían cobertura nacional, salvo la de Argentina que cubría las 31 principales aglomeraciones urbanas. De cualquier forma, dada la baja proporción de población rural en ese país (aproximadamente 10%), es probable que las reales estadísticas nacionales no difieran substancialmente de las obtenidas con datos urbanos.⁶¹ Por simplicidad, en el resto del libro nos referimos a “Argentina”, aunque los datos se refieran a la Argentina urbana.

Tabla 6.4
Encuestas de hogares de América Latina
usadas para calcular pobreza y desigualdad
Año 2010

⁵⁹ Estas encuestas en general coinciden con las usadas oficialmente por los países para medir pobreza. Algunos países (ej. Panamá) usan sus encuestas de condiciones de vida para el cómputo oficial de la pobreza, obteniendo estimaciones más esporádicas que con las encuestas de empleo/ingreso.

⁶⁰ Estos países también han implementado encuestas ECVs, aunque de forma irregular.

⁶¹ La desigualdad en Argentina no cambia significativamente cuando se comparan los resultados obtenidos a partir de la EPH con los que surgen de otras encuestas de hogares que incluyen pueblos rurales y pequeñas ciudades -ECV 1997 y 2001 e ISCA 2002- (Gasparini y Cruces, 2009). A diferencia de otros países de la región, gran parte de la población rural en Argentina está altamente integrada a la economía nacional. La comparación con Uruguay es ilustrativa. Este país, con una estructura socioeconómica similar a la existente en parte de Argentina, incorporó en 2006 las áreas rurales a su encuesta de hogares sin cambios sustanciales en sus estadísticas nacionales de desigualdad.

País	Nombre de la encuesta	Acronimo	Cobertura	Tipo de encuesta	Información sobre		Frecuencia	Continua o puntual	Tamaño muestral (c.2009)		
					Ingreso	Gasto			Hogares	Individuos	% de pob.
Argentina	Encuesta Permanente de Hogares-Continua	EPHC	Urbana	EEI	Si	No	Anual	Continua	35,956	120,642	0.30
Bolivia	Encuesta Continua de Hogares	ECH	Nacional	EEI-2	Si	Si	Anual	Continua	3,940	15,030	0.15
Brasil	Pesquisa Nacional por Amostra de Domicilios	PNAD	Nacional	EEI	Si	No	Anual	Puntual	129,333	399,387	0.21
Chile	Encuesta de Caracterización Socioeconómica Nacional	CASEN	Nacional	ECV-2	Si	No	C/3 años	Puntual	71,460	246,924	1.46
Colombia	Encuesta Continua de Hogares	ECH	Nacional	EEI-2	Si	Si	Anual	Continua	12,510	50,850	0.11
	Gran Encuesta Integrada de Hogares	GEIH	Nacional	EEI-2	Si	Si			112,102	410,910	0.91
Costa Rica	Encuesta de Hogares de Propósitos Múltiples	EHPM	Nacional	EEI	Si	No	Anual	Puntual	13,244	48,071	1.05
R. Dominicana	Encuesta Nacional de Fuerza de Trabajo	ENFT	Nacional	EEI	Si	No	Anual	Puntual	8,281	30,430	0.31
Ecuador	Encuesta de Condiciones de Vida	ECV	Nacional	ECV	Si	Si	Esporádica	Puntual	13,582	55,666	0.41
	Encuesta de Empleo, Desempleo y Subempleo	ENEMDU	Nacional	EEI	Si	No	Anual	Puntual	19,432	78,865	0.58
El Salvador	Encuesta de Hogares de Propósitos Múltiples	EHPM	Nacional	EEI-2	Si	Si	Anual	Continua	20,361	83,194	1.35
Guatemala	Encuesta Nacional de Condiciones de Vida	ENCOVI	Nacional	ECV	Si	Si	Esporádica	Continua	13,686	68,739	0.49
Honduras	Encuesta Permanente de Hogares de Propósitos Múltiples	EPHPM	Nacional	EEI	Si	No	Anual	Puntual	21,112	98,028	1.33
México	Encuesta Nacional de Ingresos y Gastos de los Hogares	ENIGH	Nacional	EEI-2	Si	Si	C/2 años	Continua	27,655	107,781	0.10
Nicaragua	Encuesta Nacional de Hogares sobre Medición de Nivel de Vida	EMNV	Nacional	ECV	Si	Si	C/4 años aprox.	Continua	6,884	36,614	0.64
Panamá	Encuesta de Hogares	EH	Nacional	EEI	Si	No	Anual	Puntual	13,386	49,573	1.44
Paraguay	Encuesta Permanente de Hogares	EPH	Nacional	EEI	Si	No	Anual	Continua	4,439	18,419	0.29
Perú	Encuesta Nacional de Hogares	ENAHO	Nacional	EEI-2	Si	Si	Anual	Continua	21,753	95,199	0.33
Uruguay	Encuesta Continua de Hogares	ECH	Nacional	EEI	Si	No	Anual	Continua	46,936	132,599	3.97
Venezuela	Encuesta de Hogares Por Muestreo	EHM	Nacional	EEI	Si	No	Anual	Continua	37,199	153,816	0.54

Fuente: elaboración propia.

Nota: EEI: encuesta de empleo e ingresos, EEI-2: EEI con información de gastos de consumo, ECV: encuesta de condiciones de vida, ECV-2: ECV sin información de gastos de consumo.

La gran mayoría de los países de Latinoamérica dispone de información distributiva con una frecuencia al menos anual, estando en algunos casos la información disponible por trimestres o semestres. Chile y México tienen encuestas de hogares cada 3 y 2 años respectivamente, pero cuentan con un fuerte sistema de encuestas laborales con el que monitorear de cerca la situación laboral. Guatemala ha realizado ECVs esporádicamente y no ha logrado consolidar la ENEI, una encuesta de empleo e ingresos. Finalmente, Nicaragua ha implementado ECVs regularmente cada espacio de 3 a 5 años.

Hay encuestas que se relevan durante todo el año, o durante muchos meses, mientras que en otras el esfuerzo de relevamiento se concentra en un mes particular. Durante los 2000s ha habido un movimiento hacia las encuestas de tipo continuo.

El tamaño muestral de las encuestas latinoamericanas es muy variable, lo que responde a diferencias en el tamaño de los países y estrategias muestrales distintas. Así la ECH de Uruguay releva al 4% de la población de esa nación, mientras que la ENIGH mexicana entrevista al 0.10% de los habitantes de ese país. La PNAD de Brasil cuenta con observaciones de más de cien mil hogares, mientras que las encuestas de Bolivia y Paraguay no superan los cinco mil hogares.

El Apéndice II de este libro amplía sobre la disponibilidad de encuestas en todos los países de la región, detallando algunas de sus características principales.

Sobre la comparabilidad

Las encuestas de hogares no son homogéneas entre países de la región, y ni siquiera en un país a través del tiempo, por lo que el tema de la comparabilidad de los resultados es de gran relevancia. En este libro seguimos la metodología de CEDLAS (ver <cedlas.econo.unlp.edu.ar> y hacemos todos los esfuerzos posibles para hacer las estadísticas comparables utilizando similares definiciones de variables, y aplicando métodos consistentes de procesamiento de los datos. Sin embargo, es justo reconocer que los problemas de comparabilidad persisten, fundamentalmente producto de las diferencias en los cuestionarios de las encuestas. Aparece acá el típico *trade-off* entre rigurosidad y cobertura que debe enfrentar todo investigador sobre temas empíricos. Es posible un análisis ambicioso que considere conjuntamente un grupo de países, pero al costo de perder precisión en los resultados. En ocasiones los problemas de comparabilidad son severos, por lo que es conveniente restringir el análisis. En otras, la relevancia del tema justifica arriesgarse a usar datos de varias fuentes, con el propósito de obtener resultados que permitan tener visiones generales u órdenes de magnitud del fenómeno estudiado. La decisión sobre si comparar o no resultados de distintas encuestas depende finalmente de las preferencias y objetivos de cada investigador.

Las encuestas de hogares tienen muchos problemas, pero aun así son la mejor fuente de información sobre las características socioeconómicas de la población. Pensamos que es conveniente evitar las dos posiciones extremas acerca de las encuestas: descartarlas o usarlas sin prudencia. Uno de los principales objetivos de este libro es promover con énfasis el uso de las encuestas de hogares de la región para el estudio de temas distributivos, pero fomentando a la vez un uso responsable de ese caudal de información.

7. Ingreso y consumo en las encuestas

Todas las encuestas de hogares de América Latina permiten construir variables de ingresos individuales y luego agregarlas a nivel del hogar y ajustar por tamaño y composición familiar. Varias encuestas de la región también permiten construir variables de gastos de consumo al nivel del hogar. Esta sección presenta los rasgos básicos de la construcción del ingreso y consumo familiar, dejando para el Apéndice III la discusión más detallada de un conjunto de problemas metodológicos específicos.

7.1. El ingreso

La definición más extendida de ingreso es la de Schanz-Haig-Simons: ingreso es el valor monetario del incremento de la capacidad de un individuo de consumir durante un período.⁶² Este criterio incluye los ingresos laborales por trabajo asalariado, en cuentapropia y como empleador o patrón, los ingresos de capital, incluyendo la renta de

⁶² Bajo esta definición el ingreso es igual al consumo del período más el incremento en la riqueza.

la tierra, los beneficios empresariales, las transferencias privadas y públicas y una estimación de la renta implícita de la vivienda propia.⁶³

Al construir un agregado de ingreso del hogar es importante prestar atención al período de referencia de las preguntas y llevar todas las respuestas a una misma unidad. En América Latina lo usual es relevar el ingreso del mes previo a la encuesta.⁶⁴ Siendo la unidad de tiempo de registro el mes, es indudable que las encuestas captan el ingreso *corriente*, con todas las desventajas que esto acarrea, discutidas al comienzo del capítulo.

Los ingresos laborales mensuales incluyen salarios y jornales por todo tipo de trabajo. Siempre que sea posible se distingue entre el trabajo asalariado, el trabajo por cuenta propia y el salario asignado al empleador. Para los dos últimos grupos la distinción entre remuneración al trabajo y al capital o a la capacidad empresarial es en un punto ambigua. Una persona que produce tejidos y los vende en la calle obtiene una ganancia que en parte es remuneración a su trabajo y en parte al pequeño capital invertido y a su capacidad empresarial. De cualquier forma, los cuestionarios de las encuestas no permiten un análisis profundo de este punto, por lo que el analista se ve forzado a aceptar las categorías pre-establecidas.

El ingreso no laboral se divide en general en tres grandes grupos: (i) la remuneración al capital, tierra y beneficios; (ii) las transferencias privadas y estatales, incluyendo las pensiones; y (iii) la renta implícita de la vivienda propia. La justificación para incluir este último ítem surge de reconocer que los hogares propietarios de una casa o apartamento no necesitan arrendar, por lo que se ahorran un gasto, pudiendo asignar esos recursos a otros fines. Esta retribución implícita constituye una parte de los ingresos de una familia, que no está canalizada en un mercado formal, pero que debe ser incluida como ingreso. En el Apéndice III se extiende este argumento y se detallan estrategias de estimación.

Es importante asegurarse que los ingresos extraordinarios no se incluyan en el cómputo del ingreso corriente, o bien se distribuyan de alguna forma en un período de tiempo largo. El mismo criterio se aplica a las ventas de activos como casas, automóviles o acciones y a regalos no frecuentes, seguros de vida, ganancias por juego y herencias. Naturalmente, es difícil para el analista conocer la frecuencia con la que se reciben estos ingresos, por lo que cierta arbitrariedad en la decisión de inclusión de cada ítem dudoso es inevitable.

Si bien ingreso es el “valor monetario” de la capacidad de consumir en un período, no significa que todos los ingresos se perciban en dinero. Muchos trabajadores reciben ingreso en especie; es decir, bienes y servicios que deben ser traducidos a pesos y adicionados al ingreso monetario de la persona. Existen tres tipos particulares de

⁶³ En Canberra Group (2001) se discute extensamente la medición del ingreso.

⁶⁴ En algunos países ciertos ingresos son reportados con una frecuencia diferente (ej. ingresos de capital o remesas), mientras que en algunos pocos casos también se reporta información de ingresos laborales sobre períodos mayores al mes (ej. ENIGH en México).

ingresos en especie de relevancia: (i) la producción de autoconsumo, (ii) la mencionada renta implícita por el uso de bienes durables propios, particularmente la vivienda y (iii) los subsidios estatales a servicios como la educación o la salud. En la práctica es usual incluir en el cómputo del ingreso, y por ende implícitamente en las estadísticas de pobreza y desigualdad, la producción de autoconsumo y la renta implícita de la vivienda propia, y es excepcional la consideración de los subsidios en especie. En el Apéndice III se extiende la discusión de la estimación de la producción de autoconsumo y la renta de la vivienda propia, mientras que el capítulo 9 trata extensamente el tema de los subsidios estatales implícitos en la provisión pública de bienes y servicios. Los gobiernos de todo el mundo proveen ciertos servicios sociales como educación, salud, vivienda, agua y saneamiento de forma gratuita o a precios subsidiados. Estas transferencias en especie constituyen ciertamente aportes al nivel de vida de la persona y deberían incluirse como parte de su ingreso. Dado que los mecanismos de imputación son controversiales, la práctica usual consiste en computar las estadísticas de pobreza y desigualdad *sin* incluir el ingreso implícito por transferencias estatales en especie, y abordar este tema en estudios especiales. Las estadísticas oficiales de pobreza de los países y aquellas calculadas por organismos internacionales y centros de investigación típicamente no incluyen el valor de la provisión pública de servicios sociales básicos como educación y salud.

El cuadro 7.1 muestra la participación de los distintos rubros incluidos usualmente en el ingreso en algunos países de la región. La relevancia de cada ítem en cada nación depende de razones económicas reales, pero también de la forma como se indaga sobre cada fuente en la encuesta. En particular, captar los ingresos de capital requiere de un detallado cuestionario y de un esfuerzo por parte de los encuestadores que pocos países llevan a cabo.

Cuadro 7.1

Estructura del ingreso

Participación de cada rubro en el ingreso individual y del hogar

País	Año	Ingresos individuales			Ingresos del hogar	
		Laborales	Capital	Transferencias	Laborales	No laborales
Argentina	2006	81.2	1.5	17.3	72.0	28.0
Bolivia	2007	81.9	3.1	15.0	76.6	23.4
Brasil	2007	76.9	3.2	20.0	71.5	28.4
Colombia	2006	86.0	2.9	11.0	85.7	14.3
R. Dominicana	2007	84.1	2.2	13.7	75.5	24.5
Ecuador	2007	85.8	2.6	11.6	80.9	19.1
Uruguay	2007	70.9	3.6	25.1	62.5	37.5

Fuente: elaboración propia en base a microdatos de encuestas de hogares.

Las encuestas de hogares latinoamericanas (y en general del mundo) captan esencialmente ingresos laborales y transferencias. En casi todos los países más del 80% de los ingresos relevados por la encuesta provienen del trabajo y más del 15% de transferencias. La relevancia de este rubro depende de la extensión de los sistemas de pensiones y de los programas de transferencias de ingreso. En Uruguay, donde el sistema de previsión social está muy extendido, el *share* de las transferencias en el total de los ingresos individuales supera el 25%. En cambio, en Colombia y Ecuador la

participación no alcanza al 12%. La participación de los ingresos de capital es exigua: apenas un 3% en promedio. Estos valores están asociados a las dificultades de las encuestas en captar personas muy ricas y a la generalizada subdeclaración de ese tipo de ingresos.⁶⁵

La participación de los ingresos laborales se mantiene alta, aun al considerar los ingresos del hogar, donde se incluye la renta implícita de la vivienda propia como ingreso no laboral. En promedio, el 75% de los ingresos familiares en las encuestas latinoamericanas provienen de fuentes laborales.

7.2. El consumo

Es usual construir un agregado de consumo identificando inicialmente dos grupos: los bienes de consumo corriente y los bienes durables. Es importante incluir todos los bienes y servicios comprados y consumidos dentro o fuera del hogar, aquellos de producción casera y los recibidos como regalos de otros hogares, transferencias del sector público o como retribución al trabajo.⁶⁶ En el caso de los bienes durables debe imputarse un valor asociado al uso del bien durante el período (ver abajo).

La siguiente tabla muestra la estructura del consumo en una muestra de países de la región, por quintiles. El grado de apertura de la información en general no coincide entre países, aunque típicamente alimentos, educación, salud, bienes durables y vivienda aparecen como rubros independientes.

Cuadro 7.2

Estructura del consumo por quintiles de consumo familiar per cápita

⁶⁵ En teoría debería incluirse en el cómputo todos los ingresos devengados, aunque algunos no estén aun realizados, criterio que es difícil llevar a la práctica, en particular con las ganancias de capital. Ninguna encuesta latinoamericana pregunta, por ejemplo, por el valor de los activos físicos y financieros que una persona posee al comienzo y al final del período de referencia para estimar la ganancia o pérdida de capital.

⁶⁶ Como se mencionó para el caso del ingreso, en la práctica no siempre es posible obtener estimaciones del consumo de propia producción, del consumo de los servicios provistos públicamente (ej. educación o salud), o del valor de uso implícito de los bienes durables propios (típicamente la vivienda).

Nicaragua, 2005						
	1	2	3	4	5	Total
Alimentación	56.0	53.7	49.7	45.3	32.3	43.9
Vivienda	11.6	11.9	12.5	13.4	20.5	15.3
Infraestructura	6.4	7.1	7.7	8.4	9.2	8.2
Salud	7.3	6.9	6.7	5.5	5.7	6.2
Educación	3.3	3.8	4.7	5.8	6.1	5.1
Equipamiento	1.8	1.9	2.9	3.6	5.1	3.6
Transporte	4.1	4.7	5.5	6.4	8.5	6.5
Otros	9.6	9.9	10.2	11.5	12.5	11.2
	100.0	100.0	100.0	100.0	100.0	100.0

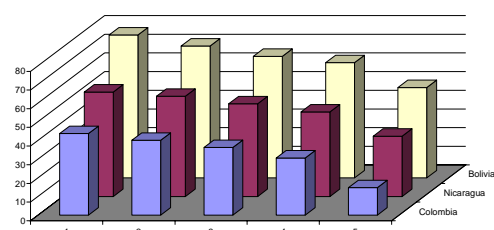
Colombia, 2005						
	1	2	3	4	5	Total
Alimentación	43.7	40.1	36.2	30.6	14.7	25.0
Vestimenta	6.5	7.2	7.4	7.3	5.3	6.3
Vivienda	20.7	20.3	21.5	21.8	19.1	20.2
Muebles	5.8	5.8	5.9	6.0	4.9	5.4
Salud	3.5	3.6	3.6	3.6	2.9	3.2
Transporte	6.3	6.7	7.7	8.4	16.0	11.9
Esparcimiento	6.0	7.2	7.5	8.3	8.6	8.1
Financieros	3.2	4.8	6.1	9.7	24.8	16.0
Otros	4.3	4.3	4.1	4.2	3.7	3.9
	100.0	100.0	100.0	99.9	100.0	100.0

Bolivia, 2005						
	1	2	3	4	5	Total
Alimentación	76.6	70.7	65.2	61.7	48.4	57.2
Educación	4.7	5.6	7.4	7.6	10.0	8.4
No alimentarios	13.7	14.7	17.4	18.6	29.5	23.3
Vivienda	5.0	9.0	10.0	12.1	12.1	11.1
Total	100.0	100.0	100.0	100.0	100.0	100.0

Fuente: elaboración propia en base a microdatos de encuestas de hogares.

Nótese el cumplimiento de la Ley de Engel en los tres países representados: la participación del consumo de alimentos en el consumo total cae a medida que se consideran quintiles más elevados (figura 7.1). El caso opuesto ocurre con los gastos en educación. Por su parte, el cuadro 7.1 indica que el *share* en vivienda resulta creciente en Bolivia y Nicaragua, pero estable en Colombia.

Figura 7.1
Participación del consumo de alimentos por quintiles



Fuente: elaboración propia en base a microdatos de encuestas de hogares.

La construcción de un agregado de consumo familiar requiere la identificación de cada ítem consumido en el período de referencia. La recolección de esta información puede ser engorrosa, dado que el número de ítems es potencialmente muy grande. De hecho, las encuestas suelen agregarlos en rubros más generales (ej. frijoles, en lugar de diferenciar por tipo de frijoles). Diferencias en el grado de agregación pueden traducirse en estimaciones del consumo, y en consecuencia de la pobreza, diferentes. Por ejemplo, Jolliffe (2001) reporta que el consumo per cápita registrado en El Salvador resultó 31% superior cuando se utilizó un cuestionario detallado largo, en comparación a un cuestionario más reducido con rubros agregados. Lanjouw y Lanjouw (2001) previenen sobre la comparación de resultados basados en encuestas aun dentro de un mismo país,

dada la práctica frecuente de modificar el diseño del cuestionario, adicionando ítems o agregándolos de forma diferente. Los autores recomiendan realizar las estimaciones sobre la base de un conjunto común de bienes en ambas encuestas.

Las encuestas de consumo suelen tener varios módulos diferenciando por unidad de tiempo de registro: de algunos bienes se reporta el consumo semanal o quincenal, de otros el mensual y de otros se apela al recuerdo del consumo anual. Estos módulos pueden ser administrados conjuntamente de una sola vez o idealmente varias veces durante el año. Naturalmente, todas las respuestas deben ser llevadas a una misma unidad de medida de tiempo (ej. mensual). En un famoso experimento en India se dividió aleatoriamente a la población en mitades, asignándoles un cuestionario donde el período de referencia de sus consumos alimentarios era una semana para un grupo y un mes para el otro (Deaton, 2000). El período de referencia más corto generó declaraciones mayores, más precisas, lo que implicó un sustancial impacto sobre la tasa estimada de pobreza; de 42.6% a 23.6% en áreas rurales (en 1998), lo que se tradujo, dado el tamaño de India en 200 millones de personas menos consideradas pobres.

A diferencia del ingreso, donde es posible definir variables a nivel individual y luego construir un agregado del hogar, en el caso del consumo se construye directamente un agregado familiar, dado que muchos gastos de consumo se realizan de forma conjunta, siendo difícil conceptualmente y en la práctica asignarlos a cada integrante. Una vez obtenido el consumo total del hogar, sí es posible realizar los ajustes por factores demográficos comentados en la sección 3. Como se discutió, el interés en el análisis distributivo en general está puesto en la persona, no en el hogar. Por esta razón, lo usual es computar pobreza o desigualdad sobre la distribución entre personas del consumo total familiar ajustado por factores demográficos.

Existen dos problemas prácticos importantes (y relacionados) en la construcción de un agregado de consumo para cada hogar: la diferenciación entre (i) consumo corriente y extraordinario y (ii) entre consumo y gasto.

Ciertos consumos son de naturaleza esporádica. Supongamos que una familia que habita en un área rural realiza en el mes en que es encuestada un viaje para asistir a una fiesta de sus parientes en la ciudad capital, lo cual trae aparejados importantes consumos en transporte, regalos y quizás alojamiento. Si la familia realiza un viaje semejante cada mes, es claro que esos gastos deben incluirse como parte de su consumo corriente. En el otro extremo, si el viaje es un hecho único en la vida, incluir su costo como parte del consumo agregado del mes de la encuesta implicaría una gruesa sobrestimación del nivel de vida de los miembros de ese hogar. Lo aconsejable en este caso es amortizar el viaje a lo largo de toda la vida de la familia y en consecuencia asignar sólo una pequeña parte de esos gastos al consumo del mes en cuestión; o por simplicidad directamente ignorar esos gastos extraordinarios. El problema en la práctica es que, como analistas, no sabemos cuán frecuente o extraordinario es el gasto reportado por la familia. La dificultad se repite con un gran número de bienes y servicios. Si una familia reporta un gasto considerable en medicamentos, ¿debemos incluirlo como parte del consumo usual de un mes típico, o ignorarlo por presumir que se trata de gastos ocasionales no

representativos? En la práctica, el analista se ve forzado a tomar decisiones arbitrarias. Dada la ausencia de un estándar internacional los institutos de estadística u organismos internacionales que construyen los agregados de consumo toman decisiones diferentes en cada país, lo que daña la comparabilidad de resultados.⁶⁷

Un segundo problema en la construcción del agregado de consumo deriva de la distinción entre consumo y gasto, hasta ahora ignorada. Pese a que en ocasiones estos dos términos se utilizan indistintamente, el gasto hace referencia a erogaciones realizadas por el hogar, mientras que el consumo está asociado al uso efectivo de bienes y servicios durante el período de referencia, independientemente de la forma como éstos fueron obtenidos. Supongamos que un hogar compra arroz cada 3 meses por valor de \$30 y lo consume de manera uniforme en los meses siguientes. Mientras que el consumo efectivo mensual es de \$10, el gasto será \$30 un mes y 0 los dos siguientes. Es claro que el concepto relevante asociado al nivel de vida es el consumo y no el gasto, por lo que se aconseja que las encuestas releven el consumo efectivo de bienes y servicios. Sin embargo, por practicidad la mayoría de las encuestas sólo preguntan por el gasto realizado por el hogar en diferentes ítems de consumo, generando sesgos. Por ejemplo, la desigualdad medida sobre la distribución del gasto resulta superior a la desigualdad medida sobre la distribución del consumo, ya que éste se distribuye más uniformemente en el tiempo.

Uno de los ejemplos más claros de la diferencia entre consumo y gasto es el de los bienes durables. El gasto se realiza en el momento de la compra, pero luego su uso se extiende por varios años durante los cuales el bien genera un flujo de utilidad. En la práctica el problema es que el consumo periódico del bien durable no es observable, a diferencia de otros bienes donde el consumo implica la desaparición del bien (como el ejemplo anterior del arroz). En este caso lo usual es asociar al consumo del bien durable con el costo de mantener el valor del stock de ese bien durante un período. Analíticamente,

$$(7.1) \quad c_{kt} = S_{kt-1}p_{kt-1} - \frac{S_{kt-1}(1-d_t) \cdot p_{kt}}{1+r_t}$$

donde c_{kt} es el consumo del bien durable k durante el período t , S es el stock del bien, p su precio de mercado, d la tasa de desgaste o depreciación y r la tasa de interés nominal. El consumo del bien durable es entonces la diferencia entre el precio del stock al inicio del período (en $t-1$) y el precio del stock resultante al final del período luego de haber tomado en cuenta el desgaste, descontado a la tasa de interés de mercado. Para implementar este cálculo es necesario conocer los precios de los bienes, la fecha de compra, e imponer alguna función de depreciación que variará con el tipo de bien.⁶⁸

⁶⁷ Deaton y Zaidi (2002) es una referencia importante para la construcción de agregados de consumo dentro del programa de LSMS.

⁶⁸ Deaton y Zaidi (2002) detallan otros cálculos cuando algunos de estos datos no son conocidos.

Para evitar cálculos que sean exageradamente afectados por la situación financiera del momento, es usual tomar la tasa de interés promedio de un período de tiempo largo.⁶⁹

Existen bienes y servicios que se consumen en el hogar pero no están asociados a un gasto monetario explícito y que deben ser incluidos en el agregado de consumo. Este grupo abarca, entre otros, a los bienes de producción doméstica consumidos en el hogar, los bienes recibidos como regalos, herencias o pagos y los servicios provistos gratuitamente por el Estado.

Para resumir, el consumo y el gasto son dos conceptos diferentes, siendo el consumo la mejor *proxy* de nivel de vida. Las encuestas de hogares suelen incluir preguntas de gasto monetario, por lo que es conveniente hacer los ajustes correspondientes para acercarse lo más posible al consumo.

7.3. Opciones en la hoja de ruta

En este capítulo hemos discutido las variables a considerar en un estudio distributivo y los pasos para su implementación empírica. Desafortunadamente, los problemas no se agotan al decidir la variable de bienestar (ingreso o consumo), los ítems a incluir y la forma de ajustar por factores demográficos. Existe un amplio conjunto de puntos metodológicos que es indispensable resolver para llegar a alguna estimación del ingreso o consumo familiar, algunos de los cuales fueron mencionados en esta sección: la estimación de la renta implícita de la vivienda propia, el cálculo de la producción de autoconsumo, el problema de la no respuesta y subdeclaración, el tratamiento a ingresos cero y valores extremos, los ajustes de precios y otros.

En este punto del libro existen dos opciones. Una es continuar profundizando en los detalles técnicos de construcción de ingreso y consumo, los cuales tienen un impacto importante sobre las estimaciones de pobreza y desigualdad. El lector que prefiere completar esta revisión metodológica puede visitar primero el Apéndice III para luego volver al capítulo 4 y continuar con el libro.

La segunda opción, para el lector más ansioso en comenzar a estudiar pobreza y desigualdad y menos ávido por detalles técnicos, es detener en este punto el examen de los problemas metodológicos de construcción del ingreso y el consumo. Esta alternativa implica asumir que se han resuelto de alguna forma los problemas mencionados arriba, y que se cuenta con una variable de ingreso o consumo que se utilizará como aproximación al bienestar individual para estudiar pobreza y desigualdad. Quien elija este camino puede seguir con la lectura del capítulo 4 y diferir la lectura del Apéndice III para más adelante.

⁶⁹ Algunas bases de datos de encuestas de hogares con información de gasto ya incluyen una estimación del valor flujo del consumo de durables (ej. Nicaragua, 2005).

Apéndice: en la práctica

En este apéndice se presentan códigos de Stata que permiten replicar los resultados presentados en diferentes cuadros a lo largo del capítulo. Las encuestas de hogares procesadas, que permiten implementar los distintos ejemplos con relativa facilidad, pueden descargarse del sitio web del libro.⁷⁰

Cociente de quintiles

El siguiente bloque de código puede utilizarse para computar el cociente de quintiles extremos presentado en el cuadro 2.2 del texto. Por su parte, la forma de computar el coeficiente de Gini se explica con detalle en el apéndice práctico del capítulo 6. La línea 4 ordena las observaciones según la variable `ipcf`, utilizando la opción `stable` para mantener el orden original de aquellas observaciones con igual valor en dicha variable. En la línea 7 se genera la variable `shrp` como la suma acumulada de la variable `pondera`, para los individuos con `ipcf` positivo; así, la última observación de dicha variable contiene la población de referencia (i.e., la suma de todos los ponderadores individuales). De hecho, en la línea 8 se computa la proporción de la población total que se acumula hasta cada observación de la encuesta. Nótese que `shrp[_N]` hace referencia a la última observación (la número `_N`) de la variable `shrp`. Las líneas 10 a 16 se emplean para construir una variable que identifica quintiles de ingreso; la variable `quintil` vale 1 para el 20% más pobre de la población, 2 para el 20% siguiente, y así sucesivamente. Las líneas 19 y 22 computan el ingreso promedio de los quintiles 1 y 5, respectivamente. Las líneas 20 y 23 almacenan esos promedios en las macros locales `media_q1` y `media_q5`. Por último, la línea 26 muestra el cociente de quintiles extremos.

```

1 * cap3-cociente-quintiles.do
2
3 * ordenar por ipcf
4 sort ipcf, stable
5
6 * computar porcentaje acumulado población
7 gen shrp = sum(pondera) if ipcf>0
8 replace shrp = shrp/shrp[_N]
9
10 * identificar quintiles de ipcf
11 gen quintil = .
12 replace quintil = 1 if shrp <= 0.2
13 replace quintil = 2 if shrp > 0.2 & shrp <= 0.4
14 replace quintil = 3 if shrp > 0.4 & shrp <= 0.6
15 replace quintil = 4 if shrp > 0.6 & shrp <= 0.8
16 replace quintil = 5 if shrp > 0.8 & shrp <= 1
17
18 * ingreso promedio quintil 1
19 summ ipcf [w=pondera] if quintil ==1
20 local media_q1=r(mean)
21 * ingreso promedio quintil 5
22 summ ipcf [w=pondera] if quintil ==5
23 local media_q5=r(mean)
24
25 * mostrar resultado
26 display "ratq51 = " `media_q5' / `media_q1'
```

⁷⁰ En particular, las encuestas procesadas contienen todas las variables que utilizamos a continuación.

En lo que resta de este apéndice, el cociente de quintiles extremos se computa varias veces. Por lo tanto, a continuación se muestra un programa que permite computar el cociente de quintiles extremos muy fácilmente; simplemente se lo invoca indicando la variable de la que se quieren computar quintiles, y opcionalmente la variable de ponderación y la condición *if*. Así, dos llamadas típicas a este programa junto con los resultados que generan son

```
. ratq51 ipcf [w=pondera]
(importance weights assumed)
media quintil 5 ipcf / media quintil 1 ipcf = 14.843565

. ratq51 ipcf [w=pondera] if ipcf>0
(importance weights assumed)
media quintil 5 ipcf / media quintil 1 ipcf = 13.43414
```

En Stata, un programa es un conjunto de instrucciones que, a diferencia de un archivo **do**, se almacena en la memoria. En general, utilizaremos programas para extender agregar capacidades del Stata agregando nuestros propios comandos. El código completo del programa se muestra a continuación. La línea 3 indica que el contenido de las líneas 4 a 52 será ejecutado cada vez que se invoque al programa `ratq51`, que pertenece a la clase `rclass`; es decir, `ratq51` es un programa que devolverá resultados en `r(resultado)`.⁷¹ La sentencia `syntax` de la línea 4 indica que este programa deberá ser invocado con una única variable como argumento, aceptando de manera opcional ponderadores y la condición *if*. El uso de ponderadores y la condición *if* son opcionales porque `iweight` e `if` aparecen entre corchetes.

La sentencia `syntax varlist(max=1)` hace que la única variable que permite especificar nuestro programa se almacene en la macro local `varlist`. Así, al invocar el programa `ratq51`, las apariciones de `'varlist'` serán reemplazadas por el nombre de la variable de la que se desea computar el cociente de quintiles extremos. La sentencia `preserve` de la línea 8 se utiliza para almacenar una copia de la base de datos; así, una vez finalizada la ejecución del programa, se recupera esta copia de la base de datos -- esto es útil si dentro del programa se eliminan observaciones de la base de datos. De hecho, la sentencia `restore` de la línea 45 recupera la copia almacenada de la base de datos. En las líneas 9 a 12 se eliminan las observaciones que no cumplen con la condición *if* especificada al invocar el programa `ratq51` -- notar que dicha condición *if* es opcional por lo que puede no especificarse.⁷² Las líneas 14-17 se emplean para asignar a la macro local `wt` el nombre de la variable de ponderación especificada al invocar el programa, o el valor 1 en caso de no utilizar ponderadores. La línea 20 ordena la base de datos de forma creciente según la variable especificada al

⁷¹ En el apéndice al capítulo 2 se explica el significado de `r(resultado)` en el marco del comando `summarize`.

⁷² En el capítulo 4 se brinda algo más de detalle respecto de la implementación de condición *if* en un programa.

invocar el programa (i.e., el contenido de la macro local `varlist`). La línea 23 se utiliza para declarar las variables temporales `shrpap` y `quintil`; así, si la base de datos contiene variables con los nombres `shrpap` y/o `quintil` no se producirá un error. Nótese que cuando se hace referencia a una variable temporal, su nombre debe aparecer entre comillas simples, las mismas que se utilizan con las macros locales.⁷³ El resto de las sentencias incluidas en el programa son similares a las presentadas en el bloque de código anterior. La principal diferencia radica en que aquí el cómputo se realiza sobre la variable cuyo nombre está contenido en la macro local `varlist`, la variable de ponderación en la macro local `wt`, y la condición `if` en la macro local `if`. Además, en este caso se utiliza un bucle para asignar quintiles de ingreso (ver líneas 29-33). Finalmente, cabe recalcar que las variables `shrpap` y `quintil` son variables temporales; por lo tanto, no serán agregadas a la encuesta de hogares.

```

1 * ratq51.do
2
3 program define ratq51, rclass
4   syntax varlist(max=1) [if] [iweight]
5
6   quietly {
7
8     preserve
9     * touse = 1 -> observación si cumple if & !=.
10    * touse = 0 -> observation no cumple if | ==.
11    marksample touse
12    keep if `touse' == 1
13
14    local wt : word 2 of `exp'
15    if "`wt'"==" " {
16      local wt = 1
17    }
18
19    * ordenar por ipcf
20    sort `varlist', stable
21
22    * variables temporales
23    tempvar shrpap quintil
24
25    * computar porcentaje acumulado población
26    gen `shrpap' = sum(`wt')
27    replace `shrpap' = `shrpap' / `shrpap'[_N]
28
29    * identificar quintiles de `varlist'
30    gen `quintil' = .
31    forvalues i = 1(1)5 {
32      replace `quintil' = `i' if `shrpap' > (`i'-1)*0.2 & `shrpap' <= `i'*0.2
33    }
34
35    * ingreso promedio quintil 1
36    summ `varlist' [`weight'`exp'] if `quintil' ==1
37    local media_q1=r(mean)
38    * ingreso promedio cuantil 5
39    summ `varlist' [`weight'`exp'] if `quintil' ==5
40    local media_q5=r(mean)
41
42    * mostrar resultado
43    local ratq51 = `media_q5' / `media_q1'
44
45    restore

```

⁷³ En realidad, el nombre de la variable temporal hace referencia a una macro local que contiene el verdadero nombre de la variable temporal; se trata de un nombre que solo Stata puede utilizar. Así, se evita que las variables temporales tengan el mismo nombre que una variable ya existente en la base de datos.

```

46  }
47
48  display as text "media quintil 5 `varlist' / media quintil 1 `varlist' = " ///
49      as result `ratq51'
50
51  return scalar ratq51 = `ratq51'
52
53 end

```

Se deja como ejercicio para el lector replicar el cuadro 2.2 pero utilizando el programa `ratq51` en lugar del bloque de código presentado más arriba. Adicionalmente, a fin de comprobar el funcionamiento del comando `syntax`, el lector puede agregar sentencias que muestren el contenido de las macros locales `varlist`, `weight`, `exp` e `if`.

El bloque de código a continuación permite identificar cuantiles de una variable cualquiera. En términos del programa `ratq51`, nos permite generar variables similares a “quintil” pero que pueden identificar quintiles, deciles, percentiles, etc. Como veremos, el programa `gcuan` se utiliza de manera intensiva en distintos capítulos del libro. Se deja como ejercicio para el lector descifrar el funcionamiento del programa. Ciertamente, los contenidos del apéndice I y del apéndice práctico del capítulo 4 serán de utilidad en esa tarea. Por lo pronto, cabe mencionar que, en este caso particular, no resulta práctico utilizar las sentencias `preserve-keep-restore` para considerar la condición `if`, debido a que nuestro objetivo es agregar a la base de datos una variable que indique el cuantil al que pertenece cada observación.

```

1  * gcuan.do
2
3  capture program drop gcuan
4  program define gcuan
5      syntax varlist(max=1) [if] [iweight], Ncuantiles(integer) Generate(namelist)
6
7      quietly {
8
9          * touse = 1 -> observación si cumple if & !=.
10         * touse = 0 -> observación no cumple if | ==.
11         marksample touse
12
13         local wt : word 2 of `exp'
14         if "`wt'"==" " {
15             local wt = 1
16         }
17
18         * variables temporales
19         tempvar myvar shrpwt popwt
20
21         * hacer copia de `varlist'
22         gen `myvar' = `varlist'
23         replace `myvar' = . if `touse' != 1
24
25         * ordenar por `varlist'
26         sort `myvar', stable
27
28         gen `popwt' = `wt'
29         replace `popwt' = 0 if `touse' != 1
30
31         * computar porcentaje acumulado población
32         gen double `shrpwt' = sum(`popwt')
33         replace `shrpwt' = `shrpwt' / `shrpwt'[_N]
34
35         * shr de la encuesta que pertenece a cada cuantil (ej 20% si ncuantiles=5)

```

```

36 local shrcuantil = 1/`ncuantiles'
37
38 * nombre variable a generar con numero de cuantil para cada observación
39 local cuantil = "`generate'"
40
41 * identificar cuantiles de `varlist'
42 gen `cuantil' = .
43 forvalues i = 1(1)`ncuantiles' {
44     replace `cuantil' = `i' if `shrpopt' > (`i'-1)*`shrcuantil' ///
45     & `shrpopt' <= `i'*`shrcuantil' & `myvar' !=.
46 }
47
48 }
49
50 * mostrar descripción cuantiles
51 tabulate `cuantil' [ `weight' `exp' ], summ(`varlist')
52
53 end

```

A modo de ejemplo, si desea generarse la variable `quintil` para identificar a qué quintil de `ipcf` pertenece cada observación, puede ejecutarse la línea

```

. gcuan ipcf [w=ponderal], n(5) g(quintil)
(importance weights assumed)

```

quintil	Summary of ingreso per capita			
	Mean	Std. Dev.	Freq.	Obs.
1	44.704461	27.8192	1992559	1992559
2	156.75712	33.171521	1994461	1994461
3	277.20735	41.995017	1993555	1993555
4	498.8373	92.16029	1992941	1992941
5	1604.412	1910.641	1995298	1995298
Total	516.5894	1025.6153	9968814	9968814

donde puede observarse que el programa `gcuan` muestra, sólo a título informativo, algunas estadísticas a partir de la variable generada.

Tamaño de los hogares

El código siguiente puede utilizarse para computar las estadísticas sobre proporción de hogares unipersonales y multipersonales presentadas en el cuadro 3.1 del texto. La línea 4 ordena la base de datos por hogar (variable `id`), ubicando en la primera observación de cada hogar al jefe de hogar – la variable `jefe` es igual a 1 para los individuos que son jefes de hogar; el comando `gsort`, a diferencia del comando `sort`, permite ordenar observaciones de mayor a menor. En la línea 5 se “marca” una única observación en cada hogar (ver variable `hh`). Es decir, la variable `hh` vale 1 para una única observación de cada hogar y cero para las demás. Luego, la variable `hh` se utiliza para computar el número de hogares en cada categoría de tamaño (líneas 7-14). El comando `tabulate` presenta la proporción de hogares en cada categoría de tamaño.

```

1 * cap3-tamano-hhd.do
2
3 * identificar una observación por hogar
4 gsort id -jefe

```

```

5 egen hh = tag(id)
6
7 gen tamaño = .
8 replace tamaño = 1 if miembros == 1 & hh==1
9 replace tamaño = 2 if miembros == 2 & hh==1
10 replace tamaño = 3 if miembros == 3 & hh==1
11 replace tamaño = 4 if miembros == 4 & hh==1
12 replace tamaño = 5 if miembros == 5 & hh==1
13 replace tamaño = 6 if miembros >= 6 & hh==1
14 tabulate tamaño [w=pondera]

```

La distribución intrahogar

El fragmento de código siguiente puede utilizarse para replicar los resultados presentados en el cuadro 3.5 del texto, que muestra cómo se modifica la desigualdad calculada a través del cociente de quintiles extremos cuando cambia la distribución del ingreso hacia el interior del hogar. Cabe recordar que la distribución del ingreso intrahogar se modifica mediante un impuesto proporcional al ingreso per cápita familiar combinado con un subsidio que sólo recibe el jefe de hogar.

La línea 4 ordena las observaciones según su código de hogar (ver variable *id*), manteniendo el orden original para los miembros de un mismo hogar. El valor de la tasa impositiva (10%) que enfrentan todos los miembros del hogar se almacena en la macro local *ty* (ver línea 7). La línea 13 se utiliza para computar la recaudación tributaria por hogar; para cada hogar se suma la variable *impuesto*, almacenando el resultado en la variable *subsidio*. El total recaudado lo recibe el jefe de hogar (línea 16). Por último, el ingreso luego de impuestos y subsidios se computa en la línea 19 (ver variable *ipcf_star*). Por construcción, este esquema de impuestos y subsidios altera la distribución del ingreso intrahogar. Las líneas de código restante se emplean para computar el cociente del ingreso promedio de los quintiles 5 y 1 como indicador de desigualdad; el código utiliza el programa *ratq51* presentado con anterioridad.

```

1 cap3-intrahogar.do
2
3 * ordenar por id
4 sort id, stable
5
6 * tasa del impuesto
7 local ty = 0.1
8
9 * impuesto al ipcf
10 gen impuesto = ipcf * `ty'
11
12 * recaudación impuesto total por hogar
13 by id: egen subsidio = sum(impuesto)
14
15 * subsidio solo lo recibe el jefe de hogar
16 replace subsidio = 0 if jefe !=1
17
18 * nuevo ipcf
19 gen ipcf_star = ipcf - impuesto + subsidio
20
21 * computar cociente quintil 5 / quintil 1
22 ratq51 ipcf_star [w=pondera] if ipcf>0

```

Empleo de ponderadores

El bloque de código que sigue puede utilizarse para construir un cuadro como el 6.1 del texto, que muestra la relación entre el ingreso per cápita familiar y el valor de la variable de ponderación. La línea 7 genera la variable `shrobs`, que contiene el porcentaje acumulado de observaciones. Como las observaciones fueron ordenadas según su `ipcf`, la última observación (la número `_N`) se refiere al individuo más rico de la encuesta. Cabe recordar que `_n` contiene el número de observación que corresponde a cada fila de la encuesta; por su parte, `_N` indica el número de observaciones de la base de datos.⁷⁴ Las líneas 9 a 15 generan la variable `quintil` que vale 1 para el 20% más pobre de la población, 2 para el 20% siguiente, y así sucesivamente; notar que, a diferencia de los cálculos anteriores, aquí incluimos a los individuos con `ipcf=0` al generar los quintiles de `ipcf`. Por último, la línea 17 computa el ingreso promedio y el factor de expansión promedio para cada quintil de `ipcf`.

```

1 * cap3-ponderadores-ing.do
2
3 * ordenar por ipcf
4 sort ipcf, stable
5
6 * computar porcentaje acumulado observaciones
7 gen shrobs = _n/_N
8
9 * identificar quintiles de ipcf
10 gen quintil = .
11 replace quintil = 1 if shrobs <= 0.2
12 replace quintil = 2 if shrobs > 0.2 & shrobs <= 0.4
13 replace quintil = 3 if shrobs > 0.4 & shrobs <= 0.6
14 replace quintil = 4 if shrobs > 0.6 & shrobs <= 0.8
15 replace quintil = 5 if shrobs > 0.8 & shrobs <= 1
16
17 table quintil, c(mean ipcf mean pondera)

```

El bloque de código siguiente calcula las tasas de pobreza con ponderadores para el total del país y para cada una de las regiones de México en 2006 (ver cuadro 6.2 del texto). En la línea 4 se asigna a la macro local `lp` el valor de la línea de pobreza de USD 2.5 por día. En las líneas 6-8 se genera la variable `pobre` que vale 1 para los individuos pobres (i.e., `ipcf < `lp'`) y cero para el resto. En las líneas 10-12 se calcula la tasa de incidencia de la pobreza como el cociente entre la suma de los ponderados de los individuos identificados como pobres y la suma de los ponderadores de toda la encuesta; es decir, la población de referencia. Las líneas 25-28 realizan el mismo cálculo para cada una de las ocho regiones que se identifican en la encuesta mexicana (véase la condición `if` en la línea 26). La sentencia

```

forvalues i = 1(1)8 {
    ...
}

```

⁷⁴ En la terminología de Stata, `_n` y `_N` son variables del sistema creadas y actualizadas por Stata de manera automática.

repite 8 veces los comandos contenidos entre las llaves; la macro local *i* va desde 1 hasta 8 a intervalos de 1.⁷⁵ Para repetir el ejercicio pero omitiendo los ponderadores, la forma más sencilla de hacerlo es agregar, al inicio, una línea de código que reemplace por uno el valor de la variable *pondera* para todas las observaciones; es decir, `replace pondera=1`.

```

1 * cap3-ponderadores-p0-mex.do
2
3 * línea de pobreza us$ 2.5 México 2006
4 local lp = 633.90918
5
6 * identificar individuos pobres
7 gen pobre = 1 if ipcf < `lp'
8 replace pobre = 0 if pobre!=1
9
10 * total país
11 summ pobre [w=pondera]
12 display r(sum)/r(sum_w)
13
14 /*
15 región
16 1 = Noroeste
17 2 = Norte
18 3 = Noreste
19 4 = Centro-Occidente
20 5 = Centro-Este
21 6 = Sur
22 7 = Oriente
23 8 = Península de Yucatán
24 */
25 forvalues i = 1(1)8 {
26     summ pobre [w=pondera] if region==`i'
27     display "H = " 100*r(sum)/r(sum_w)
28 }

```

Diseño muestral

En este apartado se muestra cómo puede considerarse la estructura muestral al momento de computar un indicador relativamente sencillo; ver cuadro 6.3 del texto del capítulo. A modo de ejemplo, se emplea la Encuesta de Hogares de Propósitos Múltiples (EHPM) de Costa Rica para el año 2006 para calcular la proporción de trabajadores empleados en las industrias alimenticia y textil. En esta aplicación, utilizaremos las sentencias de Stata agrupadas en la categoría “survey”, que permiten realizar estimaciones teniendo en cuenta el diseño muestral. En términos generales, se trata de comandos a los que se antepone el prefijo *svy*.

En primer lugar, se generan quintiles de ingreso laboral (ver variable *ila*) para los individuos ocupados y que declaran sector de empleo, empleando nuestro comando *gcuan* (ver líneas 3-4). Las líneas 6-8 generan la variable *lowtec* que vale 1 para los individuos empleados en las industrias alimenticia y textil (variable *sector*=2), 0 para los trabajadores empleados en otros sectores, y *missing* para quienes no tienen asignado un quintil de ingreso laboral.⁷⁶ Como se discutió en el capítulo, la estratificación de la

⁷⁵ En el apéndice I del libro se explica con algo más de detalle cómo pueden codificarse bucles en Stata.

⁷⁶ La variable *sector* fue generada a partir de información contenida en la encuesta original.

EHPM de Costa Rica emplea dos criterios: región y zona urbana o rural. Por lo tanto, es necesario crear la variable de estratificación que permita identificar a cuál de los estratos pertenece cada observación de la encuesta (ver línea 10). La sentencia `egen` (*extensions to generate*) extiende el comando `generate` en varias direcciones, dependiendo de qué función se utilice a la derecha del signo igual. En nuestro caso, la función `group` genera una nueva variable (`estrato` en el ejemplo) que identifica con un número a cada uno de los “grupos” diferentes en que puede dividirse la base de datos al combinar las variables que recibe como argumento. El primero de los grupos recibe el número 1, el segundo grupo recibe el número 2, y así sucesivamente.⁷⁷

En primer lugar, se ignora por completo el diseño muestral (ver líneas 13-18). Así, se utiliza el comando `ci` sin ponderadores para computar intervalos de confianza para la proporción de individuos con `lowtec=1` en cada quintil de `ila`. El resto del código emplea los comandos tipo “survey” para realizar el mismo cálculo. La línea 21 indica que el diseño muestral está definido únicamente por la existencia de ponderadores, contenidos en la variable `pondera`. Luego, la línea 22 computa el promedio de la variable `lowtec` teniendo en cuenta el diseño muestral; para ello, se antepone el prefijo `svy:` al comando `mean`. En segundo lugar, sólo se consideran los factores de expansión o ponderadores (ver líneas 20-23). En tercer lugar, se consideran los ponderadores junto con la estratificación de la muestra (ver líneas 25-28). Finalmente, se incorpora la estructura de *clusters* (ver líneas 30-33). La variable `consecu`, contenida en la encuesta original, indica a qué segmento censal pertenece cada observación. En los últimos tres casos, se computa el efecto del diseño muestral sobre las estimaciones de varianza; para ello, se emplea el comando `estat effects`.⁷⁸

```

1 * cap3-ej-marco-muestral.do
2
3 * generar quintil de ingreso laboral
4 gcuil ila [w=pondera] if ocupado==1 & sector!=., n(5) g(quintil_ila)
5
6 * lowtec=1 si empleado en industrias de baja tecnología
7 gen lowtec = 1 if sector==2 & quintil_ila !=.
8 replace lowtec = 0 if sector!=2 & quintil_ila !=.
9
10 * generar variable para indicar estrato
11 egen estrato = group(region urbano)
12
13 * 1. sin considerar el diseño muestral
14 ci lowtec if quintil_ila==1
15 ci lowtec if quintil_ila==2
16 ci lowtec if quintil_ila==3
17 ci lowtec if quintil_ila==4
18 ci lowtec if quintil_ila==5
19
20 * 2. considerando sólo factores de expansión
21 svyset [pw=pondera]
22 svy: mean lowtec, over(quintil_ila)
23 estat effects
24
25 * 3. considerando factores de expansión y estratificación

```

⁷⁷ El comando `egen` permite realizar gran cantidad de operaciones diferentes. Por lo tanto, se recomienda al lector revisar la entrada correspondiente en el manual de Stata.

⁷⁸ En el cuerpo del capítulo se explica cómo debe interpretarse dicho efecto.


```

26 svyset [pw=pondera], strata(estrato)
27 svy: mean lowtec, over(quintil_ila)
28 estat effects
29
30 * 4. considerando factores de expansión, estratificación y "clustering"
31 svyset consecu [pw=pondera], strata(estrato)
32 svy: mean lowtec, over(quintil_ila)
33 estat effects

```

Fuentes de ingreso

El bloque de código a continuación muestra cómo computar la importancia que tiene cada fuente de ingresos identificada en las encuestas de hogares (ver cuadro 7.1), a saber: laboral (variable *ila*), jubilaciones (*ijubi*), capital (*icap*), transferencias (*itran*), y otros (*ionl*). El código no presenta mayor dificultad; luego de cada sentencia `summarize` se utiliza una macro local para almacenar la suma de cada variable de ingreso. Al finalizar (ver líneas 22-27), se muestran los resultados con el comando `display` utilizando la opción `_newline` para que la participación en el ingreso total de cada fuente de ingreso aparezca en una fila diferente, a fin de facilitar la lectura.

```

1 * cap3-fuentes-ing.do
2
3 summ ila [w=pondera]
4 local ila = r(sum)
5
6 summ ijubi [w=pondera]
7 local ijubi = r(sum)
8
9 summ icap [w=pondera]
10 local icap = r(sum)
11
12 summ itran [w=pondera]
13 local itran = r(sum)
14
15 summ ionl [w=pondera]
16 local ionl = r(sum)
17
18 egen itot = rsum(ila ijubi icap itran ionl)
19 summ itot [w=pondera]
20 local itot = r(sum)
21
22 display      "shr% laboral      = " 100*`ila'/'`itot'   ///
23 _newline    "shr% jubilacion   = " 100*`ijubi'/'`itot'  ///
24 _newline    "shr% capital      = " 100*`icap'/'`itot'   ///
25 _newline    "shr% transferencia = " 100*`itran'/'`itot'  ///
26 _newline    "shr% otros        = " 100*`ionl'/'`itot'   ///
27 _newline    "shr% total        = " 100*(`ila'+`ijubi'+`icap'+`itran'+`ionl')/'`itot'

```