

Coleta de microdados para comparações internacionais *

Janes A. de Souza **

Robert Ferber ***

1. Por que microdados? 2. Meios de coleta de microdados; 3. Estudo de casos para países específicos; 4. Conclusões.

Abordar-se-ão, neste artigo vários aspectos relacionados à coleta de dados microeconômicos em países em desenvolvimento, dando-se maior atenção à América Latina, área onde os autores tiveram experiência anterior em tal tipo de levantamento.

O artigo começa discutindo a natureza dos microdados e o papel que exercem ou podem exercer no planejamento e tomada de decisão para o desenvolvimento econômico. O segundo capítulo analisa os meios de coleta de microdados, dando especial atenção aos prós e contras relacionados à crescente demanda de informações derivadas de registros administrativos como fonte de tais dados. Para ilustrar o tipo de problemas que podem ser encontrados na coleta de microdados, apresentam-se no terceiro capítulo estudos de casos de investigações que utilizaram processos

* Este artigo foi derivado de um trabalho apresentado na Conferência Internacional sobre Renda, Consumo e Preços, realizada em Hamburgo, Alemanha, em outubro de 1973 e patrocinada pelo Institut für Iberoamerika-Kund, Hamburg e The Brookings Institution, Washington.

** Do Instituto Brasileiro de Economia da Fundação Getulio Vargas, Brasil.

*** Da Universidade de Illinois, EUA.

de coleta de dados em três países latino-americanos: Brasil, Colômbia, Paraguai.

O capítulo final arrola certo número de problemas relacionados com a comparação internacional de microdados, sintetizando os resultados dos capítulos precedentes e considerando as implicações destes levantamentos para coleta e análise de dados no futuro.

1. Por que microdados?

Os chamados microdados relacionam-se com as atitudes e comportamentos de unidades de uma economia, principalmente unidades familiares e empresas comerciais e industriais. Para o setor empresarial, usualmente é a empresa ou o estabelecimento comercial ou industrial. O conceito de micro-unidades também pode aplicar-se ao setor governamental tal como nos estudos relacionados a órgãos específicos do governo municipal, estadual ou federal, embora neste aspecto a distinção entre micro e macro-unidades não seja muito nítida.

O papel dos microdados na análise econômica tem aspecto duplo: explicar tendências ou variações nos macrodados correspondentes e obter novas séries e conhecimentos sobre o comportamento econômico que não possam ser obtidos de outro modo. Isto porque as séries de macrodados, por sua própria natureza, constituem-se de várias componentes, sofrendo, portanto, diversas influências de acontecimentos econômicos e políticos.

A única forma de compreender as razões de variações, em certos setores da economia, consiste em olhar as componentes individualmente, examinando-lhes o comportamento. Isto envolve freqüentemente entrevistas com unidades econômicas individuais. Por exemplo, um aumento nas despesas agregadas (macrodados) de consumo pode ser atribuído a uma elevação repentina na compra de certos bens duráveis. Entretanto, é possível que, ao serem entrevistadas as unidades familiares individuais (microdados), revele-se que essa elevação repentina foi provocada pelas expectativas de aumentos de preços desses bens em futuro próximo. Do ponto de vista da política governamental, as informações do último tipo são de valor crucial, tanto para projeção de tendências de gastos como para qualquer decisão que se faça necessária na solução deste problema.

Os levantamentos realizados, durante vários anos, em alguns países, sobre planos empresariais de investimento em bens de capital e estoques,

ilustram um tipo de informação sobre o comportamento econômico, que somente se pode obter mediante o uso de microdados. Um ótimo exemplo são os dados sobre este assunto coletados pelo Instituto IFO de Munique, que revelam ser fontes de informações vitais tanto para projeções econométricas como para compreensão das tendências da economia na Alemanha Ocidental. Compilam-se tais dados de amostras de firmas comerciais e industriais, sendo muito úteis quando analisados por setores industriais individuais e até segundo classes de indústria.

A contrapartida no setor consumo são as investigações sobre a atitude dos consumidores, baseadas em pesquisa por amostragem de unidades de consumo, que se revelam, em certo número de países, de grande valor para a predição e explicação do desenvolvimento econômico.¹ Por outro lado, as informações obtidas, simultaneamente, sobre as razões para certos tipos de atitudes e comportamentos podem ser de inestimável valor para uma melhor compreensão de reação pessoal a respeito de diferentes políticas econômicas e os efeitos prováveis de tais políticas no comportamento futuro.

Um conjunto-chave de microdados, de certa importância no estudo do desenvolvimento, são as informações sobre padrões de consumo e renda de unidades familiares. Essas informações podem servir, por um lado, para formar estimativas mais realistas das contas nacionais de um país e, por outro, para indicar as diferenças de padrões de consumo, segundo níveis de renda e outros fatores socioeconômicos. As estimativas de elasticidade-renda e outros parâmetros derivados desses dados constituem matéria-prima na preparação de qualquer modelo de previsão ou para o planejamento do desenvolvimento econômico. Os dados de renda e poupança, que essas investigações podem fornecer constituem material básico sobre a distribuição de renda do país, fontes de renda do tipo de consumidores e disponibilidade de recursos pessoais para poupança e investimento.

Os microdados apresentam propriedades analíticas raramente encontradas em macrodados, face ao seu grau de pormenorização. Como são relacionados a unidades da atividade econômica e freqüentemente disponíveis na forma de componentes individuais, os dados podem ser rearran-

¹ Ver, por exemplo, Mueller, Eva. Ten years of consumer attitude surveys: their forecasting record. *Journal of the American Statistical Association*, n. 58, p. 899-917 Dec. 1963; Poser G. & Heckeltjen, P. *The use of anticipatory data in a quarterly econometric model of the Federal Republic of Germany's economy*. Technische Hochschule, Darmstadt, West Germany, manuscrito inédito, 1973; Shapiro, H. T. & Angevine, G. E. Consumer attitudes and buying expenditures — an analysis of the Canadian data. *Journal of Canadian Economics*, p. 230-49, 2 May 1969.

jados e cruzados de diferentes maneiras, para alcançar as finalidades de um estudo específico.

As análises podem ser efetuadas em diferentes níveis, uma vez que os microdados podem ser combinados para formar agregados, enquanto a recíproca não é verdadeira. Pode-se verificar facilmente, por esse meio, por exemplo, em que medida um aumento no preço da carne reduzir-lhe-á a demanda, mais ou menos uniformemente, em todos os grupos da população ou somente entre os grupos de mais baixa renda. Os dados agregados das contas nacionais permitem avaliar as despesas privadas de uma economia, enquanto as investigações junto a unidades familiares possibilitam medir essas despesas segundo tipos de família e, dentro de cada agrupamento familiar, segundo tipo de dispêndio.

A maior flexibilidade dos microdados também torna as comparações internacionais mais exeqüíveis, uma vez que os conceitos sobre o mesmo tópico, em diferentes países, ajustam-se uns aos outros mais facilmente.. Entretanto, as comparações internacionais não são, de nenhum modo, fáceis, envolvendo grande número de ciladas, que serão discutidas no capítulo 4.

O grande problema do uso de microdados é, em primeiro lugar, sua coleta e, por isso, a maior parte deste artigo concentra-se neste tópico. Se os dados podem ser coletados com unidade e segurança adequadas e a custos razoáveis, não somente serão disponíveis mais facilmente, como a análise e interpretação será grandemente facilitada. Por isso continuaremos, no próximo capítulo, a discutir vários meios de coleta desses dados, apresentando em complementação alguns estudos de casos nos capítulos seguintes.

2. Meios de coleta de microdados

Até o princípio da II Guerra Mundial, o levantamento de dados estatísticos em países subdesenvolvidos restringia-se à realização de censos periódicos e à estimativa, quase sempre por métodos subjetivos, de informações estatísticas globais, tais como o valor da produção agrícola e industrial, volume de produção de certos produtos, etc., uma vez que os administradores, os estatísticos e os analistas socioeconômicos contentavam-se com dados estatísticos globais para o país e, no máximo, com

informações globais sobre determinadas regiões.² Esses trabalhos não despertavam interesse geral, uma vez que a participação do Governo na atividade econômica era mínima e os usos de tais dados para fins privados, praticamente inexistentes. A qualidade das informações coletadas também era precária, pois não havia formação de pessoal qualificado e o estatístico era, em geral, “apenas um funcionário não-qualificado, encarregado de compilar informações sobre as atividades do Estado e apresentá-las sob a forma de tabelas”.³

Nos últimos 30 anos a situação mudou dramaticamente. A intervenção do Governo na atividade econômica tem crescido cada vez mais e, nos últimos 20 anos, a grande maioria dos países subdesenvolvidos vêm-se empenhando na elaboração de planos de desenvolvimento econômico, para expandir a produção e elevar os níveis de vida.

Dada essa nova situação, os dados gerais e globais, até então coletados e publicados, passaram a não mais satisfazer aos pesquisadores, cada vez mais necessitados de dados precisos e específicos. Daí o grande desenvolvimento dos processos e técnicas de coleta de microdados verificado nos últimos anos.

A coleta de microdados pode ser do tipo *cross-section*, isto é, coletando informações sobre microunidades de decisão num ponto no tempo, ou do tipo contínuo, que fornece informações referentes a pontos sucessivos no tempo.⁴ Pode ser do tipo censitário, abrangendo todo o universo de microunidades, ou por amostragem. Pode, ainda, referir-se a estudos de casos específicos, ou ser derivada de informações dadas para outros fins, como arquivos fiscais e registros administrativos.

As principais formas de coleta de microdados ora utilizadas na América Latina são as seguintes:

1. Pesquisas diretas de microunidades:
 - a) com relação a indivíduos ou famílias:
 - censos demográficos;
 - pesquisas nacionais de amostra por domicílios;
 - pesquisas sobre orçamentos familiares;

² Jaffe, A. J. Notes on developing countries and their statistics. *Review of income and Wealth*, p. 313-21, Sep. 1972.

³ Camara, Lourival. Citado em Governo da Paraíba, estatística e desenvolvimento, trabalho apresentado na XXI Conferência Nacional de Estatística, Rio de Janeiro, 1968.

⁴ Walts, H. W. Microdata: lessons from the SEO and the graduated work incentive experiment. *Annals of Economic and Social Measurement*, p. 183-91, Apr. 1972.

pesquisas específicas, de tipo qualitativo, sobre intenções e atitudes dos consumidores.

b) com relação a empresas ou firmas:

censos econômicos;

pesquisas periódicas por amostragem, sobre produção, emprego, capital empregado, etc.;

pesquisas periódicas sobre custos, níveis de tecnologia, processos de produção, etc.;

“sondagem conjuntural” sobre expectativas dos empreendedores.

2. Dados coletados para propósitos administrativos:

a) com relação a indivíduos ou famílias:

dados de formulários de imposto de renda;

dados derivados da previdência social;

dados relativos a serviços de controle de migrações;

dados derivados de folhas de alistamento militar e de alistamento eleitoral;

dados derivados de registro de nascimento;

dados de matrículas escolares.

b) com relação a empresas ou firmas:

dados de formulários de imposto de renda;

dados de impostos sobre produção industrial, sobre vendas, etc.;

dados obtidos a partir de balanços de sociedades anônimas;

dados retirados de licença para importação e exportação.

Analisar-se-á, a seguir, cada um destes tipos de investigação, inclusive, para melhor entendimento, no que diz respeito aos aspectos de sua evolução histórica.

2.1 Censos

Os *censos de população*, ao contrário dos demais levantamentos estatísticos ao nível de microunidades, vêm sendo realizados desde os períodos mais remotos da civilização. Os primeiros censos, entretanto, abrangiam apenas a contagem da população com fins bastante limitados: estavam

vinculados à coleta de impostos, a listagens eleitorais ou ao alistamento militar.⁵

Os estatísticos e economistas, em geral, não incluem os recenseamentos entre as principais fontes para a montagem de arquivos de microdados. O desenvolvimento de facilidades de computação eletrônica, em larga escala, possibilitou a montagem e a armazenagem de dados censitários a nível de microunidades, o que permite a utilização dos algarismos dos censos como o ponto de partida e elemento de base para quase todos os conjuntos de microdados coletados nos últimos anos. Isso porque, no seu conceito mais atualizado, pode-se definir o censo como processo de coleta, armazenagem e publicação de dados demográficos, econômicos e sociais, cobrindo, num período específico de tempo, todas as pessoas, domicílios, empresas, firmas e unidades políticas de um país ou região específica.⁶

Apesar de fornecer uma fotografia ampla e minuciosa da estrutura econômica e social de um país, os censos apresentam dois problemas básicos: primeiro, em vista do elevado custo, só podem ser feitos a largos períodos de tempo (geralmente 10 anos); segundo, face ao volume e diversidade de informações coligidas, têm de ser bastante gerais, não descendo a pormenores muitas vezes essenciais ao analista econômico.

2.2 Pesquisa por amostragem

Como as informações fornecidas pelos censos demográficos só são disponíveis de 10 em 10 anos e, muitas vezes, face à sua complexidade e extensão, fornecem dados com no mínimo dois a três anos de atraso, surgiram na década de 1960 as Pesquisas Nacionais de Amostra de Domicílios, investigações realizadas a curtos períodos — trimestral ou semestralmente — destinadas a fornecer, a prazos curtos e com grande rapidez, informações sobre as características básicas da população e das condições de habitação, tais como natalidade e mortalidade, fecundidade, migrações internas, flutuações da força de trabalho, emprego e desemprego, grau de instrução e saúde, padrão de vida e outras indicações do gênero.⁷ Estas informações tornaram-se tão importantes que sua coleta constitui hoje um projeto do qual participam todos os países da América Latina, o Projeto Atlântida,

⁵ Kerstenetzky, Isaac. *Tudo sobre o recenseamento*. Bloch Editores, Rio de Janeiro, 1972.

⁶ Id. *Ibid.*

⁷ Brackett J. W. & Reed R. H. *Uso analítico de datos obtenidos encuestas de hogares, trabalho apresentado no Second Symposium on Household Sample Surveys in Latin America*, Rio de Janeiro, 1970.

constituído a partir de um documento metodológico básico, elaborado pelos técnicos do Bureau do Censo dos Estados Unidos em 1967.⁸

Outro tipo de coleta de microdados que cada vez assume maior importância são as Pesquisas sobre Orçamentos Familiares.

As primeiras pesquisas sobre orçamentos familiares realizadas no mundo, no princípio do século, tinham por objetivo fornecer dados para estudos de combate à pobreza e redução das diferenças na distribuição da renda nacional do país ou região para os quais se realizavam os estudos. Por outro lado, procurava-se, a partir dos dados coletados, determinar qual o volume de gastos exigido por famílias de diferentes tamanhos, não só para sua sobrevivência, como para manter eficiência no trabalho. Entre as primeiras investigações sobre orçamentos familiares nos tempos modernos, merece especial atenção a pesquisa de orçamentos familiares de trabalhadores urbanos realizada na Inglaterra em 1904, as investigações sobre consumo de alimentos da classe trabalhadora e a pesquisa ampla sobre orçamentos familiares de assalariados, realizadas nos Estados Unidos respectivamente em 1901 e 1918/19.

Com o surgimento de problemas inflacionários em vários países, a partir do fim da I Guerra Mundial, a finalidade das investigações sobre orçamentos familiares evoluiu primordialmente para a coleta de informações básicas para montagem do sistema de "pesos" para índices do custo de vida.

Com a evolução do sistema estatístico, verificada nos últimos 30 anos, a procura de dados decorrentes de investigações sobre orçamentos familiares aumentou consideravelmente, não só devido aos problemas do custo de vida, como também por constituírem dados essenciais para medir uma série de fenômenos socioeconômicos tais como: estrutura da renda, despesa e poupança de indivíduos e famílias; estimativas de funções de demanda e avaliação da situação nutricional da população.

São três os métodos mais usuais de coleta de microdados de orçamentos familiares, todos utilizando a amostragem:

a) método de memória, que compreende o preenchimento de um questionário, abrangendo informações sobre um período dilatado de tempo. Apresenta a vantagem de maior facilidade de preenchimento e menor custo, permitindo aplicação muito mais ampla que os demais; fornecem, porém, resultados mais imprecisos;

⁸ U. S. Bureau of the Census. *Atlantida: a case study of household sample surveys*. Washington, D. C., 1967.

b) preenchimento de cadernetas familiares, pela própria unidade familiar. É de difícil aplicação em países subdesenvolvidos, por exigir grau de instrução razoável das famílias participantes;

c) acompanhamento diário de despesas e pesagem direta de alimentos por entrevistador especializado. Apresenta resultados precisos, mas é muito caro e de difícil execução, o que limita a sua aplicação a pequeno número de famílias.

As pesquisas sobre orçamentos familiares já vêm sendo feitas há vários anos em alguns países da América Latina. No Brasil, por exemplo, já se vêm realizando estudos desse tipo por diferentes órgãos, desde 1936. No entanto, somente na segunda metade da década de 60 é que foram instituídas, de maneira sistemática e comparável, em países da América Latina.

Quanto a *pesquisas qualitativas de antecipação de comportamento de consumidores*, embora já estejam bastante evoluídas em certos países como os Estados Unidos (Michigan University) e Alemanha (IFO Institut), ainda não estão difundidas na América Latina.

Com relação ao *setor empresarial*, a coleta de microdados, com exceção de dados de censos, já vem sendo realizada há muito mais tempo e com riqueza de minúcias bem maior que as investigações junto às famílias. Já se vêm fazendo investigações por amostragem, junto a empresas e firmas industriais e comerciais, com periodicidade anual e às vezes mensal, em vários países da América Latina, desde os últimos anos da década de 40, abrangendo dados como: tipo de atividade, número de empregados, volume e valor da produção, salários pagos e consumo de energia (combustíveis, energia elétrica e outros). Apenas o setor agrícola conta com menor volume de dados coletados diretamente junto às unidades produtoras. Isso porque só recentemente os investigadores demonstraram interesse maior por este setor, voltados, como estavam, para o desenvolvimento industrial, considerado por largos anos como único setor capaz de acelerar o processo de desenvolvimento dos países latino-americanos. Por outro lado, as pesquisas diretas junto a microunidades do setor agrícola são de realização muito mais difícil, face às grandes distâncias entre estabelecimentos e dificuldades de acesso.

Com relação à coleta de microdados de tipo qualitativo de *antecipação de comportamento empresarial* (sondagens conjunturais) há cerca de oito anos vêm-se desenvolvendo em países latino-americanos, sobretudo Argentina, Brasil e México. Este tipo de investigações torna-se cada vez

mais comum, pois é de execução relativamente simples e apuração rápida. O processo consiste em fazer indagações, de tipo qualitativo, sobre as intenções e planos dos empresários nos próximos meses (geralmente um trimestre) com respeito à política da empresa (ou estabelecimento) relativos à investimentos, produção, preços, mão-de-obra empregada, nível de estoques, demanda, consumo de matérias-primas, utilização da capacidade produtiva e outras.

2.3 Registros administrativos

Os microdados tanto podem ser coletados, diretamente junto a microunidades, como finalidade estatística específica, como podem ser retirados dos "registros administrativos", isto é, registros realizados por órgãos públicos ou privados, relacionados com suas atividades-fins, como cadastramento de pessoas e firmas para vários efeitos (previdência social, alistamento eleitoral e militar, registro civil de nascimento e morte, matrículas escolares e outras) ou como formulários para pagamento de impostos (imposto de renda, imposto sobre vendas, além de outros) .

Como já se disse em outra parte deste capítulo as primeiras coletas de microdados realizadas no mundo tiveram um caráter muito mais de registro administrativo do que de levantamentos estatísticos propriamente ditos. Ainda mais recentemente, no período que antecedeu às idéias desenvolvimentistas e de planejamento econômico, surgidas após a II Guerra Mundial, os pesquisadores, em vários países latino-americanos, tinham que se apoiar em informações dos registros administrativos existentes em diversas repartições do Estado, já que esses países não dispunham de serviços estatísticos organizados.

Durante vários anos, essas informações foram relegadas a segundo plano, mas, em períodos mais recentes, vêm reaparecendo como fonte de informação de enorme interesse para a mensuração de importantes aspectos socioeconômicos.

Existe o problema de escolher entre coletar microdados em pesquisas diretas ou por consulta a arquivos administrativos. Ambos os processos apresentam vantagens e desvantagens.

As principais vantagens das pesquisas diretas junto a microunidades decorrem do fato de os dados, coletados com finalidades específicas e diretas, fornecerem informações com a definição e a amplitude desejada pelos pesquisadores que vão utilizá-las. Há que considerar, porém, três

tipos de desvantagens para esse tipo de coleta. Como primeira desvantagem, surge o alto custo e, muitas vezes, em países de baixa renda, dificuldades em conseguir fontes financeiras capazes de fornecer os fundos necessários. Os governos que contam com recursos limitados e escassos estão, em geral, mais interessados em investir no aperfeiçoamento de suas fontes arrecadoras de impostos, de forma a aumentar a receita, do que gastar parte de seus recursos em levantamentos estatísticos que permitam conhecer melhor a situação socioeconômica de seus países.

A segunda desvantagem está ligada ao tempo de coleta e apuração desses dados. Mesmo contando-se com os modernos processos eletrônicos de processamento de dados, em geral gasta-se um grande período de tempo no processo de coleta, crítica e processamento de grandes volumes de microdados levantados por pesquisa direta, de forma que, quando se tornam disponíveis, já não interessam mais ao pesquisador, por desatualizados.

A terceira desvantagem diz respeito à qualidade da informação coletada. Não existe na legislação de nenhum país latino-americano, pelo que sabemos, qualquer sanção legal por fornecimento de dados imprecisos ou errados, quando a finalidade é meramente estatística. Sendo muito baixo o padrão cultural médio da população desses países, não há a consciência da utilidade de informação e nenhum esforço do informante em fornecer dados realmente fidedignos.

Quanto à coleta de microdados por meio de registros administrativos, também apresentam vantagens e desvantagens. A primeira vantagem decorre do baixo custo e da rapidez com que podem ser obtidos, uma vez que, como já são coletados e processados regularmente para outras finalidades, o custo adicional é em geral apenas o custo incorrido em preparar tabulações especiais, que melhor sirvam às finalidades da pesquisa socioeconômica. A segunda vantagem decorre da relativa segurança com que são fornecidas as informações originais, pois, como existem sanções legais algumas vezes bastante fortes por informações falsas ou erradas, os informantes são muito mais cuidadosos no preenchimento das folhas de coleta.

As desvantagens, porém, são de diversas ordens. A primeira delas é que os dados coletados com outras finalidades dificilmente fornecem informações suficientemente completas para fins de análises socioeconômicas. Por exemplo, os estudos de distribuição de renda baseados em informações dos arquivos do imposto de renda deixam de lado toda a população de baixo nível de renda e, portanto, isenta do imposto. Por outro lado, as informações quase sempre se referem à “renda tributável” e não “renda total”, deixando de considerar grande parte das rendas de capitais, não

sujeitas a impostos, e as rendas não-monetárias. As informações derivadas dos registros de previdência social não abrangem os trabalhadores que não são contribuintes ou beneficiários de previdência social que, nos países subdesenvolvidos, constituem a quase totalidade de desempregados, trabalhadores autônomos e trabalhadores agrícolas. Poder-se-iam dar vários outros exemplos deste tipo.

Outra desvantagem que pode ocorrer diz respeito a informações intencionalmente distorcidas. Como quase todos esses levantamentos estão ligados, de forma direta ou indireta, à cobrança de taxas, pode haver um “viés” deliberado no fornecimento das informações, para reduzir o pagamento de impostos.

Assim, para permitir-lhe o uso, os dados derivados de registros administrativos devem ser antes “adaptados” para preencher os requisitos da pesquisa a que se destinam. A primeira forma, a mais simples, consiste em agregar os dados primários de forma diferente da feita com propósitos não estatísticos e tabulá-los de forma conveniente. A segunda forma consistiria na complementação de informações com a junção de dados de diferentes fontes. A terceira, finalmente, seria convencer os responsáveis pela coleta das informações a adaptar os questionários, registros e arquivos, de forma a atingir tanto os objetos administrativos, quanto os estatísticos, trabalho esse nem sempre fácil, pois muitas vezes é difícil convencer a um cobrador de impostos a se preocupar (e gastar recursos e trabalho) em problemas tais como distribuição de renda ou capacidade de poupança da população.

Qualquer dos meios de coletar microdados apresenta vantagens e desvantagens. Por outro lado, nenhum conjunto simples de dados, ainda os derivados de investigações diretas, é capaz de fornecer todos os dados de que necessita o pesquisador para a maioria de seus propósitos, a não ser em casos de pesquisas restritas de objetivos muito limitados. Por esse motivo, e tendo em vista a necessidade dos usuários, não se pode falar genericamente de coleta de microdados como um conjunto de dados, obtidos por uma única investigação, seguindo um único método.

Naturalmente, quando os objetivos do analista podem ser alcançados com um conjunto único, o trabalho de coleta e tratamento dos dados fica enormemente facilitado. No entanto, quando o que se pretende obter é uma explicação completa e adequada das variações nos agregados de atividade econômica, mediante a identificação de mutações a nível de microunidades, a única maneira possível consiste em relacionar fragmentos de informações de diferentes fontes relativas a unidades econômicas obser-

vadas num certo ponto do tempo, continuando-se a levantá-los em diferentes períodos de tempo.⁹

Em tais casos, existe o problema de combinar diferentes arquivos de microdados em forma utilizável. No estágio atual de conhecimentos técnicos, os métodos de conjugação de diferentes arquivos de microdados ainda são primários e rudimentares.¹⁰ Cada caso precisa ser estudado como caso individual, devendo a solução ser específica para cada um deles. O único método válido de julgar se um arquivo sintético de microdados derivados de diferentes fontes é apropriado ou não consiste em avaliar se atinge os objetivos da pesquisa para que foi construído.

3. Estudo de casos para países específicos

Parece útil, para obter melhor conhecimento sobre a forma de condução da coleta de microdados em países menos desenvolvidos, bem como para avaliar os tipos de problemas encontrados, rever alguns estudos realizados nos últimos anos.

Este capítulo apresenta dois tipos de “estudos de casos”.

Um é uma pesquisa sobre estruturas de renda e consumo familiar em que tanto o planejamento como a execução diferem substancialmente de um país para outro, embora o objetivo primordial fosse o de realizar comparações internacionais. A natureza dessas diferenças aparece quando se descreve a maneira pela qual o estudo foi realizado em três países: Brasil, Colômbia e Paraguai.

O segundo é uma pesquisa sobre preços, onde se pôde usar planos e processos idênticos para todos os países. Por essas razões apresenta-se a descrição geral deste estudo, destacando-se as pequenas variações ocorridas de país a país. Alguns problemas específicos encontrados são salientados com referência ao Brasil.

No estudo de cada caso foi utilizado um procedimento mais ou menos padronizado. Após breve descrição do *background* a propósito do estudo particular, a apresentação enfoca os seguintes aspectos: desenho *ex-ante* da fase de coleta de dados; como a execução na prática se ajusta aos planos

⁹ Kerstenetzky, Isaac. op. cit.

¹⁰ Okner, Benjamin A. Reply and comments on Constructing a new data base from existing micro-data sets: the 1966 merge file. *Annals of Economic and Social Measurement*, p. 359-62. July 1972.

ex-ante; as razões da divergência; a avaliação da eficiência da operação e, de uma forma bastante geral, da confiabilidade dos dados. Diferem-se as implicações mais gerais, sobre a coleta e análise de dados, para as seções seguintes:

3.1 Estudos de consumo

Sob o patrocínio do Grupo ECIEL, elaboraram-se planos em 1966 e 1967 para uma investigação sobre despesas de consumo e renda nas maiores áreas urbanas dos países da ALALC (os países sul-americanos de língua espanhola e portuguesa, mais o México). A investigação cobriria residências privadas em áreas com população de 500 mil habitantes ou mais. Quando um país tinha muitas dessas áreas, como no caso do Brasil, escolheu-se uma amostra representativa delas.

O objetivo primário do estudo foi fazer comparações internacionais sobre padrões de consumo e renda. A maior parte dos países também estava interessada em usar os dados para propósitos internos. Por outro lado, muitos deles já haviam realizado estudos semelhantes no passado. Assim sendo, utilizou-se no estudo questionário igual ou similar para permitir compatibilidade. Parte por esse motivo, parte por causa das diferenças de padrões de consumo e variações sazonais existentes entre os países, não foi possível usar procedimentos integralmente padronizados nas áreas incluídas.

A apresentação a seguir revela os tipos de diferenças que surgiram e a forma como o estudo foi conduzido nos três países.

3.1.1 Brasil

A pesquisa, no Brasil, foi realizada pelo Centro de Estatística e Econometria do Instituto Brasileiro de Economia da Fundação Getúlio Vargas. A pesquisa foi planejada de forma a permitir uma investigação em quatro etapas trimestrais, compreendendo três amostras diferentes:

- a) das famílias que seriam investigadas em quatro trimestres sucessivos;
- b) das famílias que seriam perquiridas em apenas dois trimestres;
- c) das famílias que prestariam informações em apenas um trimestre.

Três motivos presidiram a escolha deste esquema de amostragem: o primeiro, de ordem financeira, colimava minimizar os custos das operações dos trabalhos de campo; o segundo, de ordem substantiva, objetivava cap-

tar as variações estacionais nos consumos de certos bens e serviços ao longo do ano da investigação; o terceiro, por razões metodológicas, visava fornecer meio de avaliar os efeitos da mortalidade e do condicionamento do painel.

O tamanho da amostra foi estabelecido, *a priori*, em função das disponibilidades orçamentárias e de experiências anteriores em pesquisa desse gênero, já realizadas pelo Instituto em 1962: mil famílias no Rio de Janeiro, 700 famílias em Porto Alegre e 700 famílias em Recife. Assim, o desenho da amostra ficou restrito à distribuição dentro dos estratos, tomando-se com variável dimensionadora a renda e tamanho da família.

O primeiro problema, com relação à seleção da amostra, constituiu-se na construção de um rol relativo à população componente do universo a ser estudado. O último censo demográfico tinha sido realizado sete anos antes, em 1960, e estava completamente desatualizado. Havia-se realizado, mais recentemente, em 1965, um censo escolar pelas secretarias estaduais de educação, para avaliar as necessidades de ampliação da rede pública de ensino primário. Este censo levantou apenas a lista dos domicílios, os endereços e tamanho das famílias. Não havia nenhuma informação sobre renda. As informações contidas nesse censo tiveram que ser complementadas pelos registros administrativos das secretarias de finanças, relativos à cobrança de impostos prediais. Usou-se como *proxi* da variável renda o valor tributado das residências incluídas no rol.

A amostra foi desenhada para eliminar ponderações, de forma que a probabilidade de sorteio de cada família fosse igual; fez-se a seleção em dois estágios.¹¹ No primeiro estágio escolheu-se, dentro de cada setor, um número fixo de famílias. Estas famílias constituíram a “amostra inicial”, que foi estratificada segundo o tamanho das famílias e o valor tributado dos domicílios. No segundo estágio selecionaram-se, entre os domicílios da “amostra inicial”, os componentes da “amostra final”.

O desenho do questionário utilizado na investigação baseou-se na experiência e modelo empregados em pesquisa anterior de 1961-1963. Houve, porém, o cuidado de adaptá-los às finalidades específicas do estudo de comparações internacionais, o que determinou uma pormenorização bastante minuciosa dos gastos familiares, resultando daí um questionário extensivo que exigiu, em média, mais de duas horas para o preenchimento.

Em face da natureza das informações a coletar e consoante a experiência adquirida na pesquisa anterior, procurou-se utilizar nas entrevistas

¹¹ Para efeito de censo, as cidades são divididas em distritos e estes em segmentos menores chamados setores censitários.

elementos do sexo feminino, preferencialmente donas de casa. Tendo em vista a necessidade de realizarem-se entrevistas nas favelas e em certos lugares de difícil acesso, empregou-se reduzido número de entrevistadores homens.

Na cidade do Rio de Janeiro, a maioria dos entrevistadores da pesquisa consistiu de professoras primárias e donas de casa, havendo ainda um pequeno número de normalistas e estudantes de sociologia.

Em Recife, a equipe de entrevistadores constituiu-se de estudantes de ciências sociais; em Porto Alegre, de estudantes de economia e ciências sociais.

A constituição da equipe de campo tornou-se problema bastante sério, pois muitos dos entrevistadores selecionados e treinados desistiram ou tiveram que ser substituídos após as primeiras entrevistas, pela complexidade do levantamento e inadaptação ao tipo de trabalho. O problema perdurou todo o tempo da pesquisa. Tomando-se por exemplo a Guanabara, vê-se que dos 100 entrevistadores selecionados e treinados inicialmente, permaneceram, durante o primeiro trimestre, 23; no segundo trimestre, o número caiu para 16; no terceiro, para 15 e no quarto, para apenas oito.

O fato de não se poder contar com uma equipe permanente e constante de entrevistadores para todo o período da investigação contribuiu em muito para alongar em demasia a duração dos trabalhos de campo, prejudicando bastante a fidedignidade dos dados.

A receptividade dos informantes, porém, foi bastante boa e a taxa de recusas bastante pequena: 13% no Rio de Janeiro, 11% em Porto Alegre e 7% em Recife. Entretanto, a taxa de recusas não se distribuiu de maneira proporcional entre os vários estratos, sendo bem maior nas famílias de rendas mais altas.

Antes de serem considerados "prontos", os questionários foram submetidos, em várias etapas do trabalho, a uma série de procedimentos de "crítica", destinados à verificação da consistência e qualidade das informações coletadas pelos entrevistadores. Paralelamente tomavam-se providências para verificar se, de fato, as famílias dos domicílios sorteados foram entrevistadas pelos enumeradores.

Dez por cento dos questionários, escolhidos aleatoriamente, após serem revistos e criticados, foram entregues aos supervisores para, numa rápida visita, repetirem as indagações sobre a composição da família, seus hábitos de consumo alimentar e rendas. O número de informações não-confirmadas foi pequeno, não sendo suficiente para prejudicar os resul-

tados. No entanto, em alguns casos, indicou a necessidade de dispensar enumeradores.

Após as várias etapas de crítica, grande número de questionários foram devolvidos ao enumerador para esclarecimentos. Na maioria das vezes foi necessário fazer nova visita à família, para sanar dúvidas e completar informações.

Somente depois de considerados satisfatórios pelos supervisores é que os questionários foram encaminhados para as etapas posteriores de codificação e perfuração.

A avaliação final da fidedignidade das informações e dos erros de amostragem só poderão ser feitas após o processamento global dos dados, o que está sendo feito na Brookings Institution. Entretanto, algumas tabulações preliminares permitem avaliar as discrepâncias entre as informações *ex-ante*, constantes do rol e os resultados efetivos da investigação.

Tabela 1

Desenho *ex-ante* da amostra e dados *ex-post* por grupos de renda e tamanho da família — Guanabara

— primeiro trimestre —
— tamanho da família —

Classes de valor tributado		Tamanho da família								Total	
		1	2	3	4	5	6	7-9	10+	Número	%
Desenho <i>ex-ante</i>											
A		8	17	21	17	16	14	18	6	117	27,9
B		2	7	7	10	7	3	3	3	42	10,0
C		3	8	12	9	4	3	6	1	46	11,0
D		4	5	6	6	9	5	4	1	40	9,6
E		3	11	12	11	6	6	5	1	55	13,1
F		2	8	10	10	9	4	4	1	48	11,5
G		5	7	7	5	6	4	3	—	37	8,8
H		1	4	3	4	4	2	5	—	23	5,5
I		—	2	3	3	2	—	1	—	11	2,6
Total		28	69	81	75	63	41	49	13	419	—
%		6,7	16,5	19,3	17,9	15,0	9,8	11,7	3,1	—	100,0
Dados <i>ex-post</i>											
Classes de renda											
1		10	4	3	3	—	2	2	—	24	5,9
2		2	6	5	2	4	1	3	—	23	5,6
3		2	15	7	11	3	4	5	1	48	11,7
4		9	16	20	19	16	11	8	1	100	24,5
5		6	10	14	14	13	6	10	4	77	18,9
6		2	12	15	14	12	7	6	5	73	17,8
7		3	5	6	7	4	3	1	1	30	7,3
8		1	2	3	3	4	4	4	—	21	5,1
9		—	—	2	6	2	2	1	—	13	3,2
Total		35	79	75	79	58	40	40	12	409	—
%		8,6	17,1	18,3	19,3	14,2	9,8	9,8	2,9	—	100,0

A tabela 1 apresenta, por exemplo, os resultados do primeiro trimestre para a Guanabara. Os números mostram claramente que os dados *ex-ante* derivados do rol construídos para efeito de cobrança de impostos apresentam concentração no primeiro estrato, o que, se fosse verdade, significaria uma pirâmide de distribuição de renda excessivamente achatada, com quase $1/3$ da população na base. Entretanto, as informações *ex-post* dos dados da pesquisa mostram forte concentração nos grupos de renda média, o que contradiz violentamente as conclusões que poderiam ser derivadas dos dados *ex-ante*.

Essa verificação esclarece um dos problemas apontados na seção anterior sobre o uso de registros administrativos para propósitos estatísticos. Como a estratificação *ex-ante* foi feita a partir do valor tributado dos imóveis, ainda que se admitisse que a correlação entre renda e qualidade da residência não fosse perfeita, o fato não seria suficiente para explicar a grande discrepância encontrada. A explicação da divergência talvez seja a de que, sendo os objetivos da avaliação a cobrança *ad-valorem* de impostos, elas estimavam o valor de locação muito abaixo do real.

3.1.2 Colômbia

Como parte do estudo ECIEL de padrões de consumo e renda na América Latina, realizou-se pesquisa em quatro das maiores cidades da Colômbia em 1967-68: Bogotá, Barranquilla, Cáli e Medelim. As quatro cidades englobavam 22% da população urbana no período. O objetivo do estudo era comparar a renda e despesas de consumos familiares nas áreas urbanas da América Latina.

As quatro cidades selecionadas da Colômbia eram bastante representativas, do ponto de vista desse objetivo.

3.1.2.1 Planejamento

Os planos e os testes preliminares para o estudo realizaram-se, em sua maior parte, durante o ano de 1966.

Os planos incluíam o desenho de um painel que continha certo número de unidades habitacionais visitadas em cada um dos quatro trimestres do ano; algumas famílias foram entrevistadas apenas duas vezes; outras apenas uma, conforme se vê na tabela seguinte. Os números representam o tamanho inicial da amostra depois dos ajustamentos feitos na primeira "onda", devido à impossibilidade de contatos ou recusa.

Subamostra	Trimestre				Total
	1	2	3	4	
1	260	225	185	160	830
2	260	225	—	—	485
3	260	—	205	—	465
4	260	—	—	170	430
5	170	—	—	—	170
6	—	170	—	—	170
7	—	—	170	—	170
8	—	—	—	170	170
Total	1 210	620	560	500	2 890

Estimou-se que seriam entrevistadas 1 720 diferentes unidades familiares a fim de obterem-se 2 890 entrevistas completas e que, descontadas as recusas iniciais e os contatos não realizados, seriam contatadas cerca de 2 015 unidades.

O propósito do desenho era permitir a avaliação dos efeitos do condicionamento e da resultante mortalidade e ajustar os dados para reduzir “vieses” não-previsíveis (embora se reconhecesse que tais ajustamentos não seriam fáceis).

Cada subamostra foi, então, selecionada sistematicamente ao acaso, dentro da amostra principal, tomada como sistema de referência.

Para assegurar a representação dos diferentes níveis de renda, utilizaram-se três estratos (alto, médio e baixo), sendo que os dados para a estratificação foram fornecidos pelo escritório municipal de planejamento. Esses dados permitiram classificar cada bairro segundo níveis médios de renda da população. A estratificação foi feita em base regional, classificando-se cada bairro de acordo com suas condições gerais.¹²

Considerou-se como unidade de amostragem a “família” definida como um conjunto de pessoas que vivem na mesma unidade habitacional e compartilham da mesma cozinha. Excluíram-se, por motivos óbvios, as pensões, definidas como unidades habitacionais com quatro hóspedes ou mais. Para efeito de análise, as famílias dividiram-se em três tipos de unidades de consumo: a) unidades primárias (membros da família que compartilham, pelo menos as despesas de alimentação, com o principal ganhador de renda); b) unidades secundárias (membros da família que

¹² Em termos de população, cerca de 65% dos bairros foram classificados como de classe baixa, 25% de classe média e 10% de classe alta.

compartilham pelo menos as despesas de alimentação uns com os outros, mas não com o chefe da família); c) unidades suplementares (membros individuais da família que compartilham pelo menos despesas de alimentação com outros membros da família, mas que não repartem outros tipos de despesas, como empregados e filhos de criação com renda própria) .

Uma família poderia constituir-se, portanto, de uma unidade primária, unidades secundárias adicionais e unidades suplementares, (sendo cada unidade numerada progressivamente a partir de 0) . Os dados foram obtidos por entrevista baseada em questionário submetido ao chefe da unidade de consumo e/ou sua esposa. Utilizaram-se três tipos diferentes de questionários: um para as unidades primárias e secundárias, outro para as unidades suplementares e um último para os entrevistadores relatarem as circunstâncias da entrevista. Foi feito um teste preliminar utilizando-se esses questionários e cadernetas de preenchimento diário. A comparação mostrou que os dados de despesas eram aproximadamente do mesmo nível, embora a caderneta diária fornecesse maiores detalhes. Isto justificou o uso dos questionários no estudo principal.

3.1.2.2 Execução

O estudo na Colômbia foi planejado e realizado pela equipe do Centro para o Desenvolvimento Econômico (CEDE) da Universidad de Los Andes, em Bogotá. Os questionários e o manual de instruções foram elaborados por esse Centro, usando protótipos desenhados pela equipe coordenadora da Brookings Institution. Os entrevistadores, contratados e supervisionados pelo CEDE, foram estudantes da Universidade, a quem se deu treinamento intensivo no uso desse tipo de questionário. Foram treinados, particularmente, antes da primeira “onda” em procedimentos de listagem, (como base para uma seleção probabilística de unidades familiares), contato e acesso às famílias incluídas na amostra, explicando-se-lhes os objetivos e a importância do estudo, o uso das várias formas de questionário e os meios de assegurar cooperação no futuro. Depois das seções de treinamento, cada entrevistador foi obrigado a realizar duas entrevistas práticas e a comparecer a uma seção final de revisão. As seções de treinamento, nas “ondas” posteriores, foram muito mais curtas e tiveram como propósito principal os processos de revisão (exceto para os novos entrevistadores contratados), já que se usavam os mesmos processos e questionários.

O trabalho de campo ajustou-se bem aos esquemas originais, sendo que as quatro "ondas" sucessivas foram realizadas, nas quatro cidades, em maio, agosto e novembro de 1967 e em maio de 1968 respectivamente. Fizeram-se extensos testes de qualidade, incluindo verificação de 10 a 20% dos questionários de cada entrevistador, que se relacionavam com várias partes da entrevista. Corrigiu-se, também, a seleção das residências da amostra. Os questionários também foram conferidos no escritório para identificar informações incompletas, inconsistências aritméticas e outras. Quando se deparava com erros e omissões maiores, devolvia-se o questionário ao entrevistador, para trabalhos posteriores.

Os dados foram codificados em Bogotá, consoante um manual de codificação elaborado pela equipe de coordenação da Brookings Institution para todos esses estudos. Depois de perfurados em cartão e verificados mecanicamente em Bogotá, os dados foram gravados em fitas magnéticas e enviadas à Brookings, onde foram submetidas a um conjunto de programas de limpeza em computador, que incluía correção para valores extremos mediante técnicas de análise de regressão.

O estágio final de purificação dos dados consistiu na análise e procura de evidência de "vieses", devidos à mortalidade da amostra ou condicionamento do painel, em correções desses "vieses" e na inserção de valores omitidos. As conferências da representatividade da amostra não mostraram nenhuma evidência de "viés" na onda inicial, tanto para a amostra total como para as subamostras. Entretanto, encontrou-se mortalidade no painel através do tempo, com perdas mais proporcionais entre famílias pequenas, principalmente as com chefes jovens e as com chefes desempregados e em famílias com níveis de renda muito alto ou muito baixo. Os efeitos de condicionamento do painel, por outro lado, foram leves.¹³ Por isso, aplicou-se uma série de ajustes para: imputar os valores necessários, fazer estimativas trimestrais de renda e despesas de cada unidade familiar por categoria; combinar as despesas em estimativas anuais quando os dados o permitissem e ponderar as informações trimestrais e anuais para corrigir a mortalidade do painel.¹⁴

Em retrospecto, o principal problema com esse estudo parece ter sido levantado nas operações de campo, especialmente com a subamostra do painel. A dimensão da mortalidade do painel foi substancialmente maior

¹³ Uma análise completa dos vieses de amostragem desses dados está contida em Murgrove, Philip. *The collection of household income and expenditure information*. Relatório mimeografado da Brookings Institution, Washington, D. C., 1972.

¹⁴ Id.

do que a prevista, devido, em parte, sem dúvida, à falta de experiência prévia da equipe de campo, como se pode ver na comparação entre as entrevistas planejadas, a amostra inicial e a amostra inicial com reposição:

	t_1	t_2	t_3	t_4	Total
Planejado	260	225	185	160	830
Amostra e subamostra inicial	183	125	86	66	460
Entrevistas com reposição	215	241	245	228	929

O tamanho inicial da amostra com o painel foi muito menor que o esperado, crescendo, como mostra a segunda linha dessa tabulação, que a mortalidade do painel pagou uma alta taxa. Somente cerca de 1/3 dos membros originais do painel foram entrevistados nas quatro ondas, quando o esperado era 60%. A reposição imediata e substancial das desistências do painel mais que compensaram as perdas do painel, como mostra a terceira linha da tabulação. Todavia, provocou alguns problemas de difícil mensuração que inquestionavelmente prejudicaram, de certo modo, a fidedignidade dos dados.

De fato, o número global de entrevista foi bem maior que o esperado (2 949 vs. 2 890), mas a experiência geral e as lacunas a serem preenchidas por entrevista de reposição mostraram que a reposição de unidades no trimestre inicial, para serem entrevistadas de novo nas subamostras do painel n_2 a n_4 , poderia ser menor do que se previu antecipadamente. Finalmente teria sido mais simples e mais eficiente atrasar as subamostras parciais de painel, n_2 a n_4 e aumentar o tamanho da subamostra do painel n_1 , e ter-se dado mais ênfase à manutenção da cooperação daqueles membros da amostra inicial. Não obstante, obteve-se um grande volume de dados muito úteis que serviram, e ainda servem, de base para muitos estudos de padrões de consumo e renda nas principais áreas urbanas da Colômbia.¹⁵

3.1.3 Paraguai

O estudo de consumo no Paraguai, como o da Colômbia, também foi realizado dentro do Programa ECIEL, com os mesmos objetivos e sob os

¹⁵ Por exemplo, Prieto, Rafael. Eucuenta de presupuestos familiares, Barranquilla, Bogotá, Cali, Medellín. Centro de Estudios sobre Desarrollo Económico, Universidad de los Andes, 1970.

mesmos critérios. Entretanto, o Paraguai é um país muito menor, com cerca de dois milhões de habitantes e com um único centro urbano importante, Assunção, com aproximadamente 300 mil pessoas. Por isso, o estudo restringiu-se a essa cidade.

3.1.3.1 Planejamento

Os planos para os estudos foram feitos em 1969 e princípios de 1970, mais tarde do que os da maior parte dos outros países. Os planos do estudo foram preparados pelo Centro Paraguayo de Estudios de Desarrollo Económico y Social em Assunção, com a colaboração da equipe de coordenação Brookings. Os planos previram uma amostra inicial de 300 unidades residenciais (o tamanho da amostra foi limitado pela disponibilidade de fundos) e um desenho modificado do painel com três subamostras iguais, sendo cada painel entrevistado duas vezes com seis meses de intervalo; cuidou-se que as variações sazonais tinham muito menor importância para Assunção. Um teste preliminar mostrou que uma indagação semestral sobre os bens duráveis mais importantes seria mais prática.

Empregaram-se as mesmas unidades de análise usadas na Colômbia. Todavia, utilizou-se apenas um tipo de questionário para todas as unidades de consumo, baseado no questionário comum desenhado pela ECIEL vários anos antes. O teste preliminar justificou este tipo de abordagem, sendo que a principal dificuldade surgiu do fato de se contar, para o teste preliminar, com mapas pobres e instruções inadequadas para a localização das unidades habitacionais selecionadas. Como na Colômbia, a seleção da amostra foi baseada no princípio da “probabilidade de área”, isto é, estratificando-se os bairros segundo suas principais características e supondo-se que o nível de renda das famílias de um mesmo bairro não difere muito.

3.1.3.2 Execução

A seleção dos entrevistadores, seu treinamento e supervisão foram idênticos aos utilizados no estudo de consumo da Colômbia. Os entrevistadores foram estudantes universitários em ciências sociais ou trabalhadores em atividades sociais. Foram submetidos a um programa similar de treinamento e ficaram sujeitos ao mesmo tipo de supervisão e verificação usado na Colômbia.

O trabalho de campo foi executado em setembro/outubro de 1970 para a "onda 1" e em março/abril de 1971 para a "onda 2".

Os resultados da amostra foram extraordinariamente bons, como mostram os dados seguintes:

	Onda 1	Onda 2
Unidades habitacionais selecionadas	303	303
Unidades não-habitacionais	30	3
Recusas	4	1
Entrevistas	269	299

Todas as 101 famílias entrevistadas na subamostra do painel na primeira onda foram também entrevistadas na segunda.

A compilação dos dados e os processos de purificação foram os mesmos utilizados nos dados de consumo da Colômbia. Desnecessário dizer que a mortalidade no painel não teve conseqüências, e que o painel foi condicionado. A construção dos registros anuais foi muito mais simples, face ao desenho simplificado do painel; o mesmo ocorrendo com as estimativas de valores dos dados omitidos.

3.2 Estudo de preços

Esta foi outra pesquisa coordenada na América Latina pela ECIEL. Diferentemente do estudo de consumo que teve objetivos nacionais e internacionais, os do estudo de preços foram quase inteiramente internacionais, para permitir comparações de preços entre os países da ALALC, o que fez com que o desenvolvimento e a execução de uma metodologia-padrão fosse muito mais facilmente aplicável.

3.2.1 Planejamento

Foi preciso obter os preços de uma larga variedade de bens e serviços para compor uma imagem representativa de suas condições. Contemplaram-se quatro classes de produtos: bens e serviços de consumo, bens de investimentos, preços dos insumos (principalmente salários e ordenados) nos serviços públicos e preços de exportação e importação. Os preços

foram coletados na capital de cada país, exceto no Equador, onde Guayaquil foi incluída junto com Quito. A primeira classe, vale dizer os bens e serviços de consumo, foi a que cobriu, por larga margem, o maior número de itens: sua amostra consistiu de 416 itens separados. Foram coletados preços relativos a três qualidades para cada item, com cerca de sete observações para cada qualidade.

A amostra de itens foi padronizada para todos os países, sendo que a seleção baseou-se parte num estudo similar realizado em 1961 pela Comissão Econômica para a América Latina (CEPAL) e parte na experiência com o estudo de consumo.

O grupo coordenador na Brookings preparou um conjunto uniforme de questionários e todos os dados foram coletados durante o mesmo período: maio de 1968. As fontes de informação para a coleta de dados de preços ao consumidor foram selecionadas com base em probabilidade estratificada, em áreas incluídas no estudo de consumo. No conjunto, utilizou-se, em cada cidade de cerca de 500 fontes de dados, alguns preços tais como de habitação. Acresce que, em certos países, foram coletados diretamente do questionário do estudo de consumo. Toda a tabulação e processamento dos dados foi feita na Brookings, usando processos padronizados.

3.2.2 Execução

Ao passo que o procedimento uniforme e os questionários padronizados ajudaram a simplificar o planejamento e, mais tarde, os estágios de tabulação e processamento de dados, o mesmo não ocorreu com igual facilidade na implementação do estágio de coleta dos dados, apesar das visitas feitas pelo grupo de coordenação e outros a cada país, para assegurar uma seleção comparável de bens e informantes e para garantir uniformidade na coleta de dados. Embora essas visitas, sem sombra de dúvida, ajudassem a promover a padronização, especialmente porque os visitantes foram usados na padronização do treinamento dos entrevistadores, a tentativa de aplicar procedimentos padronizados em 11 países com economias e produtos altamente diferentes causou dificuldades consideráveis.

O primeiro problema, apesar do treinamento, consistiu na tendência mostrada pelos entrevistadores em todos os países de selecionar fontes de informação mais modernas e de mais fácil acesso, o que pode ter introduzido um viés para algumas mercadorias no sentido de preços mais altos.

Outro problema foi a impossibilidade de obter, algumas vezes, cotações de preços para o mesmo produto, ou, até para produto similar, em todos os países. Alguns produtos, ou qualidades de produtos, não eram disponíveis; em outros casos, os controles de preços e as suspeitas por parte dos fabricantes tornou difícil, senão impossível, obter preços realistas; em outros, os produtos não puderam fornecer cotações de preços de acordo com especificações predeterminadas (o que foi especialmente verdade para maquinarias e materiais de construção) .

O resultado foi o de que, em vez das 8 700 cotações de preços planejadas para serem coletadas em cada país, o número real situou-se entre 4 mil e 5 mil. Se o período de coleta de dados tivesse sido dilatado além de maio de 1968, poder-se-ia, sem dúvida, ter coletado mais preços. Todavia, julgou-se importante que, na medida do possível, todos os preços se referissem ao mesmo mês.

Consoante os planos originais, os dados foram codificados individualmente pelos institutos e enviados, juntamente com os questionários originais, para a Brookings. Lá, os códigos foram conferidos, os erros corrigidos após contatos com os institutos, levando-se, então, os dados a perfurar e submetê-los a um processo sofisticado de limpeza e elaboração.¹⁶ Como se utilizou o mesmo processo para todos os países e como a amostragem e a coleta dos dados foram uniformes, a comparabilidade entre países foi tão satisfatória quanto possível em tais estudos, embora, de uma maneira geral, possa-se questionar a representatividade de uma cesta de bens essencialmente igual, para todos os países.

3.2.3 O caso do Brasil

O estudo de preços no Brasil cobriu, como em outros países, um grande número de bens e serviços, assim como salários de funcionários civis.

A escolha dos bens, assim como as especificações, couberam aos coordenadores e institutos participantes. As especificações foram muito minuciosas, incluindo até marcas de fabricantes, quando possível.

Na realização desta investigação o Instituto Brasileiro de Economia teve um conjunto de vantagens comparativas e economias de escala, pois conta, há 20 anos, com um setor especializado em levantamento de preços

¹⁶ Os detalhes são dados em: Salazar-Carillo, Jorge. The use of the computer in handling large price files. Trabalho apresentado na Conferência sobre o Papel do Computador em pesquisas em Ciências Econômicas e Sociais na América Latina, Guernavaca, México, out. 1961.

e cálculo de índices de preços. Assim, a equipe utilizada foi basicamente a mesma engajada há longos anos em coleta de preços, sendo que os estabelecimentos visitados já eram, em grande parte, informantes usuais do Instituto. Entretanto, face à complexidade do trabalho e à urgência com que tinha que ser feito, utilizaram-se apenas enumeradores de bom nível cultural e pertencentes a classe social elevada que, por essa razão, tinham acesso fácil a diretores de empresas. Não houve, por isto, nenhuma dificuldade na aproximação com os dirigentes das empresas informantes e na obtenção dos dados, quando existentes.

A seleção, com relação aos bens de consumo, dos estabelecimentos informantes, foi feita de forma a abranger os tipos mais representativos do setor de comercialização na cidade do Rio de Janeiro, no que diz respeito à categoria do estabelecimento e a classe socioeconômica de sua clientela. A coleta de preços de produtos de alimentação foi feita em nove estabelecimentos varejistas, sendo três supermercados de rede, três grandes armazéns de rede e três pequenos armazéns de bairro. Dos três estabelecimentos de cada tipo visitado, um estava localizado em zona predominantemente da classe alta, outro de classe média e outro de classe baixa.

Para artigos de vestuário, produtos eletrodomésticos e outros, selecionaram-se seis lojas, sendo duas grandes lojas de departamento, em cadeia, duas em zonas predominantemente de classe média-alta e classe alta e duas em zonas predominantemente de classe média-baixa e classe baixa.

A grande dificuldade encontrada na coleta de preços de bens de consumo foi a identificação dos produtos e a ligação entre as especificações incluídas no questionário e os tipos de produtos efetivamente encontrados no comércio brasileiro. Entre estas dificuldades cabe mencionar:

- a) especificação de produtos que foi tão pormenorizada que nem ainda os vendedores eram capazes de reconhecer exatamente qual a mercadoria referida. Exemplo: ternos de homens feitos com tecidos mistos com especificação precisa, em percentagem, das diferentes fibras naturais e sintéticas que a compunham;
- b) inexistência, no mercado brasileiro, sobretudo no Rio de Janeiro, de produtos só usados em clima frio;
- c) especificação de mercadorias que já caíram em desuso: conjunto de *ban-lon* para mulher, meia de algodão para homem;
- d) especificação de artigos só encontrados, no Brasil, em lojas de alto luxo. Por exemplo, o tipo de *lingerie* especificado só era encontrado nas *boutiques* mais sofisticadas;

e) tamanho e unidade de medida diferentes das geralmente encontradas no Brasil. Por exemplo, os lençóis encontrados no comércio brasileiro eram bem menores que os especificados; a largura dos tecidos era sempre diferente (maior ou menor). Por outro lado, as frutas no Brasil são vendidas, em geral, por dúzia e o preço pedido referia-se a quilo. Também o tamanho e peso das embalagens de alimentos enlatados nunca coincidiam com os brasileiros.

Existem ainda uma série de problemas, tais como produto de alimentação inexistentes no Brasil e cortes de carne diferentes dos utilizados.

As ponderações para a cesta de bens de consumo derivou da pesquisa sobre orçamentos familiares já citada anteriormente. Fizeram-se, apenas, duas pequenas adaptações: estimativa de um aluguel imputado às residências próprias e exclusão, para efeito do cálculo das ponderações, das famílias compreendidas nos dois decis de renda superior.

As questões relacionadas a maquinaria e equipamento e bens de consumo duráveis serão examinadas em conjunto, pois são da mesma espécie.

Da mesma forma que a lista de bens de consumo, a lista e as especificações das máquinas, equipamentos e bens de consumo durável foram enviadas pela coordenação.

A coleta dos preços desses bens foi realizada junto aos representantes comerciais dos grandes fabricantes e às maiores firmas importadoras. Em geral selecionaram-se, para cada produto, cinco empresas (algumas vezes o número de fabricantes e/ou importadores existentes não atingia esse número), que forneceram os preços dos produtos constantes de toda sua linha de produção. Assim, por exemplo, para motores elétricos de corrente alternada foram selecionados os quatro únicos fabricantes do Brasil, que forneceram preços para 46 especificações diferentes.

Os problemas de adaptação de especificações destes produtos foram muito maiores que para os bens de consumo. Embora o Brasil ainda importe quantidade considerável de máquinas e equipamentos, a maior parte destes produtos utilizados no Brasil são de produção nacional. Acresce que a legislação brasileira cria entraves que praticamente impedem a importação de mercadorias com sucedâneos próximos produzidos no País. Desta forma, em cerca de 80% dos produtos especificados na cesta de maquinaria e equipamentos e de bens de consumo durável só foram encontrados os de fabricação nacional, com marcas e outras especificações diferentes das constantes da lista distribuída pelo coordenador. Por isso, foi

necessário fazer, para cada produto, uma coleta bastante exaustiva de preços de produtos brasileiros com as características principais mais ou menos próximas à lista básica, de forma a permitir que os preços finais utilizados fossem imputados pelo método de regressão.

As ponderações para compor a cesta constituída por estes bens foram calculadas a partir das estatísticas sobre valor da produção nacional e de importações.

A pesquisa sobre salários no Serviço Público não suscitou maior problema, pois existiam informações bastante minuciosas sobre o assunto, nos registros administrativos do Departamento de Administração do Pessoal Civil.

Na pesquisa sobre preços de construção, os tipos de problemas foram inteiramente diferentes dos encontrados para os outros bens. A coleta de preços das componentes dos quatro tipos de construção selecionados (casas residenciais, edifícios de apartamentos, edifícios comerciais e rodovias) não apresentou nenhuma dificuldade, pois as especificações de componentes e materiais correspondia, exatamente, aos padrões usados no Brasil.

O grande problema consistiu em compor um sistema de ponderações para permitir agregar os quatro tipos de construção. As ponderações deveriam ser derivadas das informações sobre "investimentos", nas Contas Nacionais. Todavia, as estimativas de investimentos, no Brasil, são calculadas de forma global de acordo com o consumo aparente de insumos, discriminando apenas investimentos públicos e privados. Por isto, foi preciso estimar, ainda grosseiramente, a importância relativa de cada um dos tipos de construção na composição do fluxo de investimentos nacionais. Para isto, recorreu-se a dados derivados, tanto de fontes administrativas, como estatísticas: balanços da União, estados, municípios, autarquias, fundações e empresas governamentais, valor das construções segundo tipos, estimadas pelo Instituto Brasileiro de Geografia e Estatística para cidades com mais de 50 mil habitantes e projeções dos dados do censo de 1960, sobre construções rurais.

4. Conclusões

Baseado no material anteriormente analisado e em considerações mais gerais, parece apropriado sumariar nesse item de conclusões algumas das "ciladas" envolvidas nas comparações internacionais quando se usam microdados e sugerir o tipo de medidas que podem ser consideradas para

combatê-las, futuramente. Na opinião dos autores merecem consideração especial as seguintes:

1. Uma "cilada" que se deve observar desde logo refere-se à possibilidade de os estudos que visam levantamentos de dados para comparações internacionais poderem ser distorcidos por interesses nacionais. Se os dados pertinentes são relevantes para tomadas de decisões políticas a nível nacional e, especialmente se dados similares foram coletados para este propósito no passado, cada Instituto estará fortemente pressionado para modificar o desenho do estudo, de forma a obter dados comparáveis com os coletados no passado, em vez de dados comparáveis com aqueles a serem coletados em outros países. A tentativa de influenciar um país para que altere os quesitos, a fim de atender às finalidades de um estudo internacional, sacrificando a comparabilidade dos dados coletados para seus próprios interesses, dificilmente será bem sucedida e, provavelmente, causará problemas no futuro. Uma solução mais satisfatória consiste em desenvolver um desenho de pesquisa e um formato de questionário que possa satisfazer, com modificações apropriadas, ambas as necessidades, ainda que envolva maiores custos.

2. Diferenças estruturais de consumo e produção tornam o planejamento do estudo muito difícil, quando os países envolvidos se encontram em diferentes latitudes e têm costumes diferentes. Comparações relacionadas com a América Latina revelam as grandes diferenças de consumo associadas a condições climáticas, costumes e tradições. Bens essenciais para a população de certos países — roupas de inverno, por exemplo — não existem em outros países. Os hábitos de alimentação são também muito diferentes. Por exemplo, no Brasil, arroz e farinha de mandioca são alimentos básicos muito mais importantes que pão e batatas, e a carne de aves ainda é considerada bem superior à carne bovina.

A estrutura de produção também torna as comparações difíceis, e algumas vezes, impossíveis. Em alguns países, como Argentina, Brasil e México, o setor industrial é basicamente composto de grandes empresas em setores tais como aço, produtos petroquímicos, automóveis e navios. Em outros países, como a Bolívia e Equador, a atividade industrial básica consiste principalmente em processamento primário de produtos agrícolas e artesanatos de tipo familiar. As comparações do setor industrial nestes dois tipos de países podem levar a conclusões absurdas.

3. O capítulo anterior já ilustrou as dificuldades de padronização de unidades. As unidades de medidas, tamanhos e tipos de embalagens, qua-

lidade, marcas e tipos de produtos podem diferir inteiramente de um país para outro, de sorte que não é possível usar especificações rigorosas. A maneira de contornar o problema é muito difícil. Em certos casos pode-se relacionar medidas equivalentes. Em outros, seria preferível usar especificações mais abertas e relacionar os produtos e as categorias a seus usos ou tipos de consumo, do que procurar uma réplica exata em cada país. Ainda em outros casos, seria possível obter equivalências estatísticas pelo método de regressão, apesar de o método envolver necessariamente pressuposições de inter-relações similares entre produtos em diferentes países, o que pode não ser válido.

4. No caso dos registros administrativos, a qualidade de um tipo particular de um dado pode variar substancialmente de país a país. Algumas vezes é impossível obter os dados necessários, porque são mantidos em caráter confidencial por certos órgãos. Ainda em outras situações, os dados podem ser distorcidos intencionalmente por motivos políticos.

Por tudo isso, a menos que os dados tenham sido avaliados e considerados válidos no passado, seria prudente comparar os dados derivados dos registros administrativos com informações de outras fontes. A comparação citada anteriormente na descrição do estudo de consumo no Brasil é um bom exemplo: apesar de os dados de renda da pesquisa não poderem ser considerados totalmente precisos, servem, sem sombra de dúvida, como um bom indicador para avaliar a adequabilidade geral da variável *proxi* dos registros administrativos, usada para estratificação de renda.

5. Diferenças no tipo de material amostrado e outros recursos para pesquisas, de um país para outro, pode ter um efeito substancial na qualidade da investigação obtida para diferentes países. Se um país tem uma amostra básica atualizada a respeito de suas áreas urbanas, enquanto outro possui apenas material censitário com oito ou 10 anos de idade, pode-se esperar diferenças substanciais na representatividade das amostras obtidas para os dois países. Alocar recursos extras para obter melhor material de amostragens no último país — uma solução óbvia — nem sempre é possível, tanto do ponto de vista do pessoal quanto do de recursos.

6. Outra “cilada” decorre do fato de que os resultados *ex-post* de uma operação de pesquisa podem ser muito diferentes dos derivados dos planos *ex-anté*. Essas diferenças parecem variar substancialmente de um país a outro. Também são substanciais as diferenças na qualidade do trabalho de campo. Não é necessariamente verdade dizer que é a organização com

melhor equipe de trabalho de campo que fará o melhor trabalho. Porque a equipe de campo poderá não ter experiência naquele tipo específico de pesquisa. Também, a qualidade da supervisão é uma questão-chave e ela pode variar em demasia, de organização a organização. Dentro de certos limites, pode-se alcançar a padronização, desenvolvendo-se processos de treinamento e materiais uniformes e uniformizando-se a supervisão e as práticas de verificação. Resta o fato, entretanto, de que cada organização procura fazer algumas coisas à sua própria maneira. Apenas através de comunicação constante é que se encontra alguma semelhança de padronização.

A codificação e outras operações de elaboração de dados também estão sujeitas a variação considerável entre países, ainda quando se fornecem códigos comuns e convenções padronizadas de codificação. Idealmente a padronização da elaboração dos dados poderia ser alcançada se fosse feita integralmente em um único local. Na prática isto é geralmente impossível, podendo também ser indesejável do ponto de vista de campo. A única solução tanto para a elaboração dos dados quanto para as operações de campo consiste em prover recursos adequados para a equipe de coordenação visitar regularmente cada país, enquanto o estudo estiver em andamento e conferir os procedimentos que estão sendo usados. Confiar em malas postais, telegramas ou telefones é dificilmente exequível em países menos desenvolvidos, particularmente porque uma organização pode desconhecer que está fazendo coisas diferentes de outros países até que seja notificada, durante o correr de uma visita.

7. O transcurso de um estudo pode ser interrompido inclusive por convulsões políticas. É raramente possível conduzir um mesmo estudo, ao mesmo tempo, em vários países (o estudo de preços da ECIEL foi uma exceção), sendo difícil fixar um calendário preestabelecido.

8. Atualmente, todo país praticamente impõe certas restrições ao mercado internacional e/ou taxas de câmbio de tal forma que qualquer comparação de microdados, em termos de "valores monetários" é necessariamente distorcida por essas restrições. Ainda em países onde não existem restrições governamentais na taxa de câmbio, ela está sujeita a influência de vários fatores, de tal forma que a correlação entre este valor e o poder de compra é pequeno. Entre os fatores de influência podemos citar o luxo de bens e serviços, a composição dos preços relativos, a estabilidade política, as restrições governamentais à entrada ou saída de moeda estrangeira, fluxo de viagens de turismo, equilíbrio ou deficit do balanço de pagamentos e, naturalmente, restrições governamentais sobre a taxa de

câmbio. Para citar apenas um exemplo, em fevereiro, quando existe um aumento considerável no número de turistas no Brasil, particularmente americanos, a taxa de câmbio no mercado paralelo cai em relação ao cruzeiro; em julho, quando existe o fluxo oposto (do Brasil para os Estados Unidos da América e Europa) a taxa de câmbio aumenta, em relação ao cruzeiro.

9. Como consideração final nunca é demais salientar que a coleta de microdados para comparações internacionais envolve constante esforço de coordenação, esforço esse que não pode ser dirigido de um local fixo. A definição de conceitos e procedimentos uniformes no estágio *ex-ante* é um pré-requisito para o bom sucesso de tais operações. Todavia, não existe nenhuma garantia de que se alcançarão dados comparáveis. Para obtê-los, devem ser realizadas supervisões e conferências constantes, para garantir que os trabalhos de pesquisas no campo e o manuseio do questionário sejam realizados de maneira similar em todos os países. Paradoxalmente, porém, uma padronização rígida é autodestrutiva se não for realística no que entende com as variáveis em estudo.

É mister haver flexibilidade para compensar as diferenças de padrões de vida e produção existentes nos vários países.

LEI DO ENSAIO E DO ERRO

MUITO CARA PARA SER OBEDECIDA

Você não precisa perder tempo para atingir seus objetivos. **Conjuntura Econômica** oferece mensalmente análises completas e objetivas do comportamento da economia nacional. A situação de sua empresa fica clara. As surpresas são eliminadas, através de prospecções baseadas em dados fidedignos e análises seguras. Você fica sabendo qual o caminho a seguir. Sem perder tempo e sem perder dinheiro.

Leia **Conjuntura Econômica** e deixe os passatempos para os fins de semana.