

PESQUISA DE UM MÉTODO OBJETIVO PARA QUANTIFICAR A EXPORTAÇÃO DE MERCADORIAS PELAS VIAS INTERNAS (1)

Salomão Schattan

Neste trabalho, é descrita a pesquisa realizada na Divisão de Estatísticas Econômicas - Coordenadoria de Análise de Dados - Secretaria de Economia e Planejamento, a partir de 1976, em busca de um esquema operacional que permitisse a elaboração das estatísticas de exportação de mercadorias, do Estado de São Paulo para as demais Unidades da Federação.

A importância destas estatísticas reside no fato de embasarem as análises econômicas regionais, permitindo maior precisão ao planejamento econômico regional.

Dado o grande número de operações que constituem o processo do "comércio por vias internas" do Estado de São Paulo, cerca de um e meio milhão por mês, era forçoso adotar o método de amostragem para a elaboração das estimativas necessárias e possíveis.

Em consequência, o problema inicial consistiu na seleção das técnicas de amostragem que melhor se aplicavam.

Assim, as operações para as distintas Unidades da Federação foram consideradas e tratadas como populações independentes e escolheu-se a nota fiscal, que caracteriza uma operação de venda, como unidade de amostragem.

Razões de ordem teórica e prática levaram à adoção da amostragem como estratificação simples para as treze Unidades da Federação cujo número de notas fiscais é pequeno, enquanto para as doze unidades restantes a técnica adotada foi a de subamostragem ou amostragem dupla com estratificação.

Como "levantamento piloto" para o dimensionamento de uma amostra inicial, foram utilizados os resultados do levantamento de abril de 1974. Os dados desse levantamento permitiram estimar que uma amostra de 50.000 notas fiscais seria suficiente para estimar os totais das oito classes de mercadorias da Nomenclatura Brasileira de Mercadorias para cada mês e Estado, com um erro padrão não superior a dez por cento.

São descritos os problemas práticos de manipulação da grande massa de notas fiscais e sua preparação para o cálculo em computador.

(1) Liberado para publicação em 16/06/1981.

A seguir são descritos os dois esquemas da amostragem, sendo analisados os problemas surgidos em sua aplicação.

Os estimadores, apropriados às duas técnicas de amostragem adotadas, são apresentados de forma detalhada.

O estudo termina com a apresentação de métodos estatísticos para otimizar as amostras nos Estados.

Esta otimização consiste na determinação da conveniência de se adotar a subamostragem; na determinação dos limites dos estratos; e finalmente no compromisso das variâncias das diferentes mercadorias, pois que se trata da amostragem de características múltiplas.

Dado o caráter dinâmico da economia do Estado, espera-se que seu comércio com os demais Estados seja bastante mutável, tanto no que se refere às mercadorias exportadas como às quantidades correspondentes, o que indicaria a conveniência de se efetuar os cálculos relativos à otimização paralelamente às estimativas dos totais para que se possa efetuar as mudanças necessárias na estrutura da amostra, tão logo se verifique uma baixa inaceitável na eficiência da amostra.

1 - INTRODUÇÃO

As distintas regiões do Brasil, constituídas pelas Unidades da Federação, diferem muito entre si, tanto no que se refere à natureza das atividades econômicas quanto à intensidade dessas atividades.

As regiões se complementam e funcionam como economias abertas, pois não existem entre elas barreiras políticas e são quase nulas as barreiras institucionais e fiscais.

As diferenças regionais e a liberdade de circulação fazem com que o comércio entre as Unidades da Federação tenha importância econômica fundamental, e as "Estatísticas do Comércio por Vias Internas" devem responder à necessidade de informações sobre a natureza e o volume deste comércio.

Os cálculos da renda regional e mesmo a análise da economia de cada região exigem o conhecimento das transações inter-regionais. É óbvio que para se estimar o consumo total e o consumo per capita de um grupo de artigos em determinado Estado, torna-se indispensável conhecer, além da produção do próprio Estado, o saldo, positivo ou negativo, do comércio destes artigos com os demais Estados.

Apenas para ilustrar, tome-se como exemplo a estimação do consumo de proteínas e calorias pela população do Estado de São Paulo que se constitui em problema econômico e social importante. Mesmo conhecendo com a exatidão desejada a produção dos principais artigos agrícolas do Estado, como efetivamente se conhece, não se pode estimar com igual exatidão o consumo de sua população,

porque as fronteiras do Estado são atravessadas, permanentemente, por volumosos fluxos de produtos agrícolas nos dois sentidos. A inexistência de estatísticas fidedignas do "comércio por vias internas" impede que se faça o balanço desse comércio, impedindo igualmente a determinação dos volumes efetivamente disponíveis para o consumo interno, o que torna incompleta e insegura sua análise.

Esta mesma indefinição existe para todas as análises onde seja necessário o conhecimento do balanço do comércio inter-regional. Daí a importância da elaboração deste tipo de estatística.

O reconhecimento da importância das "Estatísticas do Comércio por Vias Internas" não é recente, pois é sabido que os responsáveis pelas estatísticas em nível nacional se preocupam de há muito com este problema, tendo sido definido um esquema em que os órgãos de estatística dos Estados apurariam suas exportações por Estado de destino e por mercadoria (3 dígitos NBM).

O conjunto dessas estatísticas permitiria a elaboração de matrizes do comércio inter-regional para produtos ou grupos de produtos, em que se visualizaria a origem e o destino desta parcela da produção nacional transacionada entre os Estados.

A implantação do esquema foi precedida de uma legislação federal que obriga as firmas industriais e comerciais, que enviam mercadorias para fora de seu Estado, a entregar mensalmente, numa agência da Fundação IBGE, cópia das notas fiscais emitidas em função das operações que implicam envio de mercadorias para outros Estados. De forma que, neste esquema, a Fundação IBGE se encarrega da coleta das informações. Estas informações na forma de Notas Fiscais ou listagem de computador são entregues aos órgãos de estatística nos Estados para apuração e publicação das "Estatísticas do Comércio por Vias Internas".

Infelizmente, o esquema assim organizado não tem produzido os resultados esperados. Mesmo para os anos dos Censos Econômicos, não foi possível completar as matrizes de fluxos, porque um certo número de Estados não consegue apurar o material que recebe. Entre eles, encontram-se justamente os mais industrializados, que participam com a maior parcela das exportações por vias internas.

No Estado de São Paulo, os poucos dados produzidos até o momento mostram-se frágeis quando se analisa em profundidade sua consistência interna. A última tentativa para implantar um esquema permanente de apuração das exportações por vias internas data de 1974. A metodologia de trabalho adotada e posteriormente abandonada no início de 1976 só permitiu o preparo do material referente aos 6 primeiros meses de 1974, e as apurações feitas pela PRODESP se referem aos meses de janeiro, fevereiro, março e abril. Técnicos da

Fundação João Pinheiro analisaram os dados de exportação de janeiro a março de 1974 para Minas Gerais, os quais apóntaram alguns erros, cuja origem não foi possível determinar, mas que põem em dúvida a fidedignidade dos resultados alcançados.

Existe, pois, um impasse representado pela necessidade dessas estatísticas do comércio por vias internas e pela impossibilidade de obtê-las com os meios postos em prática até então. Para resolver este impasse, tornou-se necessário desenvolver novas técnicas que permitissem a apuração regular dessas estatísticas e que, além disso, tornassem possível comprovar a fidedignidade dos resultados obtidos.

2 - O PROBLEMA

A frustração das inúmeras tentativas de apuração das exportações feitas a partir do Estado de São Paulo, para os demais Estados da Federação, resulta da inadequação dos métodos de trabalho aplicados durante essas tentativas, pois problemas tão ou mais complexos têm sido resolvidos na esfera da administração pública.

A determinação dos métodos apropriados de trabalho só poderia resultar da pesquisa e desenvolvimento de novas técnicas. A pesquisa destas técnicas teve início em abril de 1976.

O grande número de notas fiscais recebido mensalmente pela Fundação IBGE e a limitação de recursos afastam a possibilidade de fazer seu Censo apurando todo material, com o que se impõe a aplicação do método de amostragem.

Para se ter uma idéia da magnitude e complexidade do problema, basta lembrar que mesmo não considerando a apuração por mercadorias, que seria o ideal, mas sim a de agrupamentos homogêneos de mercadorias, definidos por três dígitos da Nomenclatura Brasileira de Mercadorias (NBM), em número de 392, as quais conjugadas com as dez classes definidas para a natureza da operação, as sete classes de via de transportes e os 25 Estados de destino, proporcionam um total de 686.000 estatísticas possíveis de valor e outras tantas estimativas de peso.

Se for considerado que a Fundação IBGE recebe mensalmente cerca de 1,2 milhão de notas fiscais, observa-se que há, em média, duas notas fiscais por parâmetro possível, o que evidentemente não permitiria a amostragem.

Na verdade, a população de notas fiscais está longe de se distribuir uniformemente, seja entre os Estados ou entre qualquer outra das subdivisões definidas.

O Rio de Janeiro, que hoje inclui a antiga Guanabara, de um lado e Fernando de Noronha, de outro, constituem, respectivamente,

o maior e o menor cliente de São Paulo. Entre as oito classes de mercadorias em que se agrupam os 392 códigos NBM, são as classes 2, 4, 6 e 7 que concentram o maior número de notas e de valor exportado. Nas demais classificações predominam as operações de venda e o transporte por rodovia.

Portanto, deve-se levar em conta esta distribuição da população para definir uma amostra capaz de estimar, com precisão pré-fixada, o peso e o valor de cada mercadoria, segundo as características definidas no esquema de publicação proposto pela Fundação IBGE.

Uma vez definida a amostra será necessário organizar a população de forma a permitir a retirada e o preparo eficiente da amostra nos prazos pré-fixados.

3 - POPULAÇÃO E UNIDADES DE AMOSTRAGEM

O universo a estudar e mensurar neste trabalho é constituído pelo conjunto de transações econômicas, que envolvem a exportação de mercadorias de um Estado para as demais Unidades da Federação, através das chamadas vias internas de transporte.

O conjunto de notas fiscais emitidas em função desse comércio constitui o "universo estatístico" ou "população", e a nota fiscal que representa uma transação será a sua unidade de amostragem.

Para produzir as "Estatísticas do Comércio Por Vias Internas" a Fundação IBGE coleta cópia das notas fiscais emitidas por ocasião do registro das transações que envolvem despacho de mercadorias para outros Estados.

A população de notas fiscais não é a única possível para o estudo do comércio por vias internas. Uma alternativa seria a população de firmas que exportam mercadorias para os demais Estados. As Secretarias da Fazenda, nos Estados, podem relacionar essas firmas porque têm necessidade de escriturar a compensação do ICM, e esta relação pode constituir um "sistema de referência" para amostragem.

3.1 - A População e Subpopulação

O esquema de trabalho adotado em 1974 definiu uma só população, consistindo de todas as notas entregues em um mês às Agências do IBGE no Estado de São Paulo.

Desta população única, era extraída uma amostra preliminar encaminhada ao Departamento de Estatística do Estado. Depois de estratificada a amostra preliminar, procedia-se à subamostragem.

A exigência de apresentar os resultados por Estado de destino

tornava necessária a pós-estratificação da subamostra, segundo os 25 Estados de destino.

Calculou-se que, com este esquema de amostragem, a estimativa do valor total das mercadorias exportadas para Fernando de Noronha no mês de abril de 1974, com erro padrão de amostragem não superior a 20%, só poderia ser conseguida com uma amostra de 250.000 notas fiscais. A amostra deste mês foi de 38.000 notas fiscais e os resultados apresentados por mercadorias, o que dá bem idéia da imprecisão dessas estimativas.

Neste caso extremo, tanto o tamanho da amostra, como os erros de amostragem das estimativas eram inaceitáveis, indicando assim a necessidade de se adotar esquema mais eficiente de amostragem.

O exame dos resultados da apuração de abril de 1974 mostrou a conveniência de separar desde o início a população de notas em subpopulações, segundo o Estado de destino.

As estimativas dos totais e dos componentes da variância correspondentes às 8 classes de mercadorias NBM em cada Estado, feitas no levantamento de abril, indicaram que adotando o critério de Neyman na alocação da amostra nos estratos e o esquema de Cochran na determinação do tamanho da amostra para variáveis múltiplas, seria necessária uma amostra global de 50.000 notas fiscais para estimar os totais de cada uma das 8 classes NBM em cada Estado de destino, com erro padrão em porcentagem não superior a 10%.

Evidentemente, para estimar os totais por mercadoria (3 dígitos NBM) com a mesma exatidão, seria necessária amostra muito maior, que entretanto não pôde ser dimensionada devido à fragilidade das estimativas dos componentes da variância nos levantamentos disponíveis.

3.2 - Unidade de Amostragem - Nota Fiscal

Apesar das tentativas feitas e não obstante existir um modelo básico de nota fiscal definido em Lei, ainda não foi possível uniformizar o formato das notas fiscais. Cada firma define seu formato de acordo com suas conveniências e, por vezes, fantasia.

O máximo que se conseguiu é que todas elas incluíssem os seguintes elementos:

- a) nome, endereço, CGC da firma emitente;
- b) data da emissão;
- c) nome e endereço do destinatário;
- d) natureza da operação;

- e) via de transporte;
- f) nome da mercadoria e código;
- g) peso de cada mercadoria constante da nota fiscal;
- h) valor total do ICM da nota, já incluído no valor da mercadoria;
- i) peso total bruto e líquido do total de mercadorias da nota; e
- j) transportadora, nome e endereço.

Destas informações, só o peso é por vezes omitido nas notas, sendo igualmente problema o preenchimento, nem sempre claro e perfeitamente legível.

3.3 - Descrição das Informações Registradas para Processamento

A seguir, são enumeradas e descritas as informações extraídas das notas fiscais, para processamento.

3.3.1 - Origem da mercadoria exportada

Numa análise em nível nacional, será suficiente conhecer os Estados de origem e destino de cada mercadoria, ou grupo de mercadorias.

Entretanto, para o Estado de São Paulo, por exemplo, não bastará saber a quantidade global de uma mercadoria exportada para cada uma das Unidades da Federação. É igualmente importante saber o Município ou Região Administrativa do Estado que exportou.

A fim de atender a esta exigência, foi decidido codificar o Município onde foi emitida a Nota Fiscal.

3.3.2 - Estado de destino

Adotou-se a codificação correspondente usada pela Fundação IBGE.

3.3.3 - Natureza da operação

Há operações que estão, e outras que não estão, sujeitas ao pagamento de impostos. Por esta razão, as firmas são obrigadas a especificar a natureza da operação.

São 10 os tipos de operações que, para os objetivos do presente trabalho, podem ser resumidos a:

– Vendas: códigos (01), (02), (03) e (07)

	Código Fazendário
Vendas a contribuintes para industrialização e/ou comercialização	(01)
Vendas a contribuintes para uso e/ou consumo próprio	(02)
Vendas a não contribuintes	(03)
Remessa para venda fora do estabelecimento	(07)

– Transferências: códigos (04), (05), (06) e (08)

	Código Fazendário
Transferência para industrialização e/ou comercialização	(04)
Transferência para uso e/ou consumo próprio	(05)
Remessa para industrialização para outro estabelecimento	(06)
Retorno de industrialização por outro estabelecimento	(08)

– Outras operações: códigos (09) e (10)

	Código Fazendário
Devoluções	(09)
Outras saídas não especificadas	(10)

3.3.4 - Via de transporte

As firmas especificam sete vias de transporte.

Considerando que a apuração será feita por amostragem, não tem sentido manter subdivisões que são pouco freqüentes, pois neste caso os elevados erros de amostragem reduzirão a utilidade dessas estimativas.

Por esta razão constituíram-se duas classes, a das remessas por rodovia, cuja freqüência é de aproximadamente 90%, reunindo as demais em uma só classe. Assim têm-se:

a) rodovia; e

b) aérea, postal, ferroviária, marítima, fluvial, saídas não especificadas.

3.3.5 - Peso da mercadoria

Esta informação pode aparecer de duas maneiras, que são equivalentes quando a nota fiscal se refere a uma única mercadoria.

No corpo principal da nota fiscal, há um espaço reservado para o peso da mercadoria, vindo a seguir as demais informações.

Na parte inferior, reservada a informações sobre a transportadora e a forma de transporte, deve ser consignado o peso líquido e bruto das mercadorias constantes da nota fiscal. Esta informação geralmente consta da nota fiscal, enquanto a informação de peso no corpo da nota fiscal é menos freqüente.

Considerando que o peso e o valor são as variáveis a serem estimadas para cada mercadoria e cada Estado de destino, faz-se necessário estimar o peso, quando o mesmo não consta da nota fiscal.

Para isso, organizou-se uma tabela do valor médio, por unidade de pesos das mercadorias incluídas em cada um dos códigos TAB, nos 3 últimos meses de operação, para servir como referência na estimativa do peso, a partir do valor da mercadoria consignado na nota fiscal, e que nunca deixa de ser preenchido.

Este procedimento não é tranqüilo, porque em um mesmo código podem estar incluídas mercadorias com valores específicos muito diferentes entre si e então o peso médio pode não representar adequadamente todas as mercadorias aí incluídas. Entretanto, esta é a única solução encontrada até o momento para este problema.

3.3.6 - Códigos das mercadorias

As publicações das estatísticas do comércio por vias internas adotam os códigos e seguem a estrutura da Nomenclatura Brasileira de Mercadorias (NBM). Ela contém as classes de 1 a 9, sendo a classe 3 vazia. São definidos aí 392 grupos distintos de mercadorias identificadas por 3 dígitos.

Esta mesma nomenclatura era adotada pelos órgãos fazendários federais e estaduais. Posteriormente, o Ministério da Fazenda desenvolveu e passou a adotar a Tarifa Aduaneira do Brasil (TAB), em substituição à NBM. O novo código é compatível com os códigos de mercadorias utilizadas no comércio internacional.

As mercadorias na TAB são definidas por 8 dígitos. Ela contém 21 seções, 1.091 agrupamentos de mercadorias definidas por 4 dígitos, e 7.500 agrupamentos definidos por 6 dígitos.

A TAB é hoje adotada por todos os órgãos fazendários do Brasil. Em especial, as alíquotas do IPI são definidas para os códigos

TAB.

As indústrias cobram do comprador o IPI e a porcentagem do imposto sobre o valor da mercadoria é definida para os diferentes códigos TAB. As notas fiscais emitidas pelas indústrias contêm, além do nome da mercadoria, o código TAB.

O IPI é cobrado uma única vez e, portanto, as firmas comerciais não estão sujeitas à cobrança desse tributo. Entretanto, quando essas firmas adotam código para as mercadorias, ou é um código interno, ou então, TAB.

O importante é que a maioria das notas fiscais contêm código, mas em nenhuma circunstância se trata de código NBM.

Na sistemática anterior, todas as mercadorias das notas fiscais da amostra recebiam o código NBM correspondente, porque os resultados são apresentados segundo o código NBM. Esta codificação se constituía na operação mais difícil e morosa de toda preparação para os cálculos.

A fim de facilitar este trabalho de codificação das mercadorias da amostra adotou-se o código TAB de 4 dígitos. Desta forma, a busca se limitava àquelas mercadorias sem código; para as demais o problema estava resolvido.

Note-se que a correspondência entre os 1091 códigos TAB e os 392 NBM não é perfeita, no sentido de que, em alguns casos, parte das mercadorias de um mesmo código TAB-4 dígitos corresponde a 1 código NBM-3 dígitos, e o restante das mercadorias a outro código NBM. Este problema de correspondência não ficaria perfeitamente resolvido, mesmo que se adotassem 6 dígitos da TAB, o que levaria a ampliar para 7.500 o número de códigos TAB.

Adotando o código TAB de 4 dígitos, a correspondência é feita para aquele código NBM que contém a maior parte das mercadorias incluídas no código TAB.

Isto envolve arbitrariedade em alguns casos e leva à mudança na definição de alguns códigos NBM, porém, uma vez definida a correspondência, será perfeitamente válido fazer a análise econômica do comportamento de suas estimativas ao longo do tempo.

Em verdade, não há nada de sacrossanto nos atuais agrupamentos de mercadorias definidos na NBM, que impeça de mudá-la; ao contrário, a operacionalização do trabalho torna essas mudanças muito convenientes, sendo mínimas as desvantagens que delas advêm.

De qualquer forma, a codificação TAB 4 dígitos com uma tabela de correspondência para NBM 3 dígitos simplifica enormemente o trabalho de codificação da amostra, tornando-o exequível com os

recursos disponíveis. Esta nova técnica traz ainda a vantagem de permitir que no final se possa produzir duas publicações, uma de acordo com a sistemática NBM, para atender à necessidade de compatibilização a nível nacional, e outra TAB, compatível com a análise econômica de dados fiscais relativos à indústria em geral.

3.3.7 - Valor

A nota fiscal contém o valor de cada mercadoria em cruzeiro, e se admite que ele corresponde sempre ao verdadeiro valor da transação.

Na prática comercial, podem ocorrer descontos sobre o valor de tabela. Quando isso ocorre, aparece o valor de tabela, o desconto correspondente e o valor líquido. Em uma mesma nota fiscal, podem ocorrer descontos diferentes para distintas mercadorias.

De qualquer forma, como se está interessado em estimar a verdadeira importância recebida pelo exportador por sua mercadoria, o valor líquido é transcrito para a planilha.

Este valor inclui o ICM, mas não inclui o IPI.

3.3.8 - Imposto sobre circulação de mercadorias (ICM)

O total do ICM efetivamente recolhido para as mercadorias constantes da nota fiscal aparece e é transcrito na planilha.

O total assim transcrito já está reduzido das isenções e abatimentos legais.

3.4 - Análise da Consistência dos Dados Transcritos

O ideal seria proceder à digitação dos dados, a partir da própria nota fiscal amostrada. Entretanto, a diversidade de tamanho, forma e disposição das informações na nota fiscal tornam a operação morosa.

Por esta razão, procede-se à transposição da informação da nota fiscal para planilhas, a partir das quais os dados são digitados.

Os elementos da amostra correspondentes a um estrato de um Estado de destino são colocados em uma única pasta, em cuja capa estão apostas as identificações correspondentes a Mês - Unidade da Federação-Estrato. Estes dados e mais o código do digitador e do codificador da pasta são duplicados, automaticamente, em todos os cartões:

Uma linha da planilha é constituída de 80 colunas contendo as seguintes informações:

Campo	Colunas
1) Código do digitador	1 - 2
2) Código do codificador	3 - 4
3) Mês	5 - 6
4) Unidade da Federação	7 - 8
5) Estrato	9

A seguir, na mesma linha, têm-se dois itens ou notas fiscais.

Campo	1º item colunas	2º item colunas
6) Cidade emissora	10 - 12	45 - 47
7) Natureza da operação	13 - 14	48 - 49
8) Via de transporte	15	50
9) Peso	16 - 20	51 - 55
10) Unidade de peso	21	56
11) Código de mercadoria	22 - 25	57 - 60
12) Valor Líquido	26 - 33	61 - 68
13) Valor do ICM	34 - 40	69 - 75
14) Número de ordem	41 - 44	76 - 79

Uma vez terminada a digitação e a conferência dos dados, é feita a análise de sua consistência, através de um programa de computador.

Os campos de 1 a 8 devem ser preenchidos com determinados códigos. Toda vez que o computador lê em um desses campos um código diferente daqueles possíveis, trata-se de erro e então é emitida uma mensagem de erro.

No caso do campo número 5, que indica o estrato, sabe-se que cada um deles corresponde a um intervalo de valor que é confrontado com o valor lido no campo 12. Toda vez que o valor lido no campo 12 não estiver compreendido no intervalo correspondente ao estrato indicado no campo 5 é emitida uma mensagem de erro.

O campo 9 (peso) é constituído de 5 colunas. A unidade de peso consignada para uma mercadoria depende da própria mercado-

ria, isto é, de seu valor específico. Se se tratar de ouro ela será pesada em gramas, enquanto o ferro é consignado em toneladas.

Dispondo de cinco colunas para registrar o peso, foi-se obrigado a criar um campo de unidade de peso, onde vai o código da unidade (grama, quilograma, tonelada, etc), coluna 21.

A experiência no manuseio de notas fiscais ensinou que a quantidade de notas fiscais sem informação sobre o peso da mercadoria não é negligenciável. A fim de sanar este inconveniente, é utilizada informação dos meses anteriores. Para cada código TAB, é calculado o peso por cruzeiro, nas notas com informação, e com elas se calcula a média dos três últimos meses. Este "peso específico médio" é multiplicado pelo "valor líquido" da nota fiscal, quando não existe informação de peso, e este produto fornece uma "estimativa" do peso da mercadoria em questão, a qual é utilizada como substituto da informação inexistente.

Considerando que cada código TAB de 4 dígitos contém diversas mercadorias e que, portanto, o "peso específico" calculado não corresponde nunca à mercadoria em questão, conclui-se que a estimativa de peso feita por seu intermédio pode estar longe da realidade.

Entretanto, o processo é válido se for considerado que o objetivo não é estimar o peso da mercadoria de uma nota fiscal, mas sim o peso médio da mercadoria de um determinado código TAB-4 dígitos, em um conjunto de notas fiscais que não informaram o peso. Espera-se que as mercadorias das notas fiscais sem informação de peso representem adequadamente as mercadorias incluídas no código em questão.

Pretende-se investigar a possibilidade de utilizar este "peso específico médio" e um intervalo de confiança adequado para analisar os "pesos específicos" das mercadorias de cada nota, com o objetivo de verificar a validade dos dados. Toda vez que o "peso específico da nota" estiver fora do intervalo de confiança do "peso específico médio" haverá emissão de mensagem de erro do computador.

O campo 10, unidade de peso, é analisado na consistência, e toda vez que aparece um código inválido é emitida mensagem de erro pelo computador.

O campo 11, código da mercadoria, tem quatro colunas, podendo aparecer nele 10 mil números distintos. Entretanto, só existem 1.091 códigos válidos. Toda vez que aparecer neste campo um código diferente destes 1.091 válidos, será emitida mensagem de erro.

O campo 12 contém 8 colunas e, portanto, o valor líquido de mercadoria pode chegar a 99 milhões de cruzeiros, o que no momento é perfeitamente suficiente. Juntamente com o peso da mercadoria,

ele permite o cálculo do peso por cruzeiro, que poderá ser utilizado na análise de consistência.

De acordo com a natureza da operação e a mercadoria, existe um percentual máximo e mínimo de ICM. Estes dados devem ser utilizados para a construção de uma tabela de código TAB com os respectivos intervalos percentuais de ICM, que permitirá verificar se o ICM consignado está no intervalo permitido, emitindo mensagem de erro toda vez que este valor estiver fora do intervalo.

As notas fiscais que constituem a amostra do estrato de determinada Unidade da Federação são numeradas em ordem crescente e sucessiva de 1 a n_j sendo n_j fornecido "a priori". Será emitida mensagem de erro toda vez que haja falta, duplicação de números e números superiores a n_j .

3.5 - Tratamento dos Erros Detectados

Os erros encontrados podem ter distintas naturezas:

- a) a nota fiscal amostrada pertence efetivamente a uma outra Unidade da Federação;
- b) erro de estratificação;
- c) erro de perfuração; e
- d) erro de codificação.

A correção dos erros de perfuração e codificação detectados é feita substituindo no arquivo de dados o registro errado pelo registro corrigido. Os cálculos finais são feitos após a substituição de todos os registros errados.

Há dois outros tipos de erro que podem ser detectados imediatamente após a seleção da amostra, por ocasião da codificação.

Trata-se de notas fiscais fora do estrato e notas fiscais pertencentes a outras Unidades da Federação. É praticamente impossível eliminar totalmente este tipo de erro. Entretanto, todos os esforços devem ser feitos para manter a proporção destes erros dentro de um limite tolerável.

Depois da seleção da amostra, nada válido poderá ser feito para remediar os inconvenientes advindos dos erros de estratificação. Se bem que este tipo de erro não produza viés nas estimativas de total, as correspondentes estimativas da variância ficam inflacionadas, podendo provocar grande redução da eficiência da amostragem em relação a uma amostra irrestrita aleatória.

O mais grave ocorre quando unidades de amostragem dos estratos maiores são encontradas nos estratos menores, porque além de serem grandes as diferenças entre estes valores e a média do estrato, são em geral grandes os fatores de elevação dos estratos pequenos, e a soma de quadrados fica inflacionada pelo produto do quadrado de ambos esses fatores.

Os erros de classificação entre as Unidades da Federação provêm de confusão de siglas como de Rondônia e Roraima, ou simplesmente por extravio inadvertido. Sua correção pode ser feita levando em conta o fato de a amostra informar sobre a verdadeira estrutura da população. A nota fiscal encontrada na amostra é indicação da existência de outras notas fiscais da mesma natureza, erradamente classificadas e misturadas na população. O número dessas notas erradamente classificadas na população é estimado pelo fator de elevação do estrato em que a nota foi encontrada. Pode ocorrer que, além do Estado, o estrato também esteja errado.

Seja i o Estado certo e k o Estado errado.

Seja j o estrato certo e ℓ o estrato errado.

Então N_{ij} é o tamanho inicial da população do Estado i e o estrato j , ao qual a nota fiscal efetivamente pertence, e $N_{k\ell}$ o tamanho inicial da população do Estado k e o estrato ℓ , onde a nota fiscal foi encontrada, e

n_{ij} e $n_{k\ell}$ tamanho inicial das respectivas amostras.

Feita a transposição da nota para sua posição certa, será necessário corrigir o tamanho da população e da amostra. Adotando o critério de integrar na amostra o elemento extraviado, o tamanho da população e da amostra será dado por:

$$N_{ij} \text{ passa a } N'_{ij} = N_{ij} + g_{k\ell}; \quad n_{ij} \text{ passa a } n'_{ij} = n_{ij} + 1;$$

$$N_{k\ell} \text{ passa a } N'_{k\ell} = N_{k\ell} - g_{k\ell}; \quad n_{k\ell} \text{ passa a } n'_{k\ell} = n_{k\ell} - 1;$$

$$g_{ij} \text{ passa a } g'_{ij} = \frac{N'_{ij}}{n'_{ij}} = \frac{N_{ij} + g_{k\ell}}{n_{ij} + 1}; \text{ e}$$

$$g_{k\ell} \text{ passa a } g'_{k\ell} = \frac{N_{k\ell}}{n'_{k\ell}} = \frac{N_{k\ell} - g_{k\ell}}{n_{k\ell} - 1}.$$

4 - ESQUEMA DE AMOSTRAGEM

Foi dito anteriormente que o esquema de amostragem adotado na tentativa de 1974 não era satisfatório e, em conseqüência, dever-se-ia adotar outro esquema mais eficiente.

A ineficiência apontada resultava da adoção de uma população única, da qual era retirada a amostra preliminar. Nestas condições, a fração de amostragem tendia a se manter constante para os diferentes Estados.

Como conseqüência, a amostra preliminar resultante, nos estratos de diversos Estados, era menor que o tamanho da amostra necessária para estimar os totais das 8 classes NBM, com a exatidão desejável.

Esta constatação levou, naturalmente, a dividir a população global em subpopulações, o que permitiu a adoção de esquemas de amostragem distintos, conforme o tamanho da subpopulação, ou seja, o número de notas fiscais em cada Estado de destino.

4.1 - Amostragem Preliminar - O Material Entregue pelo IBGE

Em conseqüência, os Estados foram separados em dois grupos: o primeiro contém 13 Estados com pequeno volume de transações, dos quais a Fundação IBGE não mais retira amostra preliminar; e o segundo grupo com 12 Estados, nos quais são separadas inicialmente as notas fiscais cujo valor líquido é igual ou superior a Cr\$100.000,00 (cem mil cruzeiros). Das notas fiscais de valor líquido inferior a Cr\$100.000,00, é extraída amostra sistemática de dez por cento.

De forma que o material recebido da Fundação IBGE se constitui das seguintes partes:

- a) a totalidade das notas fiscais relativas aos 13 Estados "pequenos";
- b) a totalidade das notas fiscais cujo valor líquido é igual ou superior a Cr\$100.000,00 nos 12 Estados "grandes";
- c) dez por cento das notas fiscais cujo valor líquido é inferior a Cr\$100.000,00 nos 12 Estados "grandes"; e
- d) listagens de cerca de 200 firmas, cuja contabilidade é mecanizada, e que substituem a obrigação de entregar cópia das notas fiscais pelas listagens resumidas das mesmas.

Três considerações devem ser feitas a respeito do material recebido:

- a) apesar de se constituir em um grande volume de documentos, não se tem segurança que o mesmo inclua, efetivamente, todas as exportações paulistas por vias internas. Isto é, não se sabe se esta é a população completa. Nos meses de janeiro a março de 1977, conseguiu-se identificar mais de 25 mil firmas que realizaram operações de exportação por vias internas. Uma primeira verificação deste universo poderia ser feita confrontando esta lista com as

informações fornecidas pelas firmas à Secretaria da Fazenda para efeito de compensação do ICM, com as demais Unidades da Federação. É possível que haja outras formas de identificar as firmas omissas;

- b) ao trabalhar com as notas fiscais, constatou-se que há inclusão indevida de notas fiscais referentes ao comércio exterior, notas de operações financeiras, como transferência de créditos de incentivos a exportações, havendo mesmo certo número de duplicações. Com o desenvolvimento dos trabalhos, tanto as omissões como as duplicações e as inclusões indevidas serão detectadas e eliminadas no momento oportuno, utilizando-se métodos que precisam ser desenvolvidos; e
- c) as listagens se originaram de solicitação das firmas, cuja contabilidade é processada em equipamento eletrônico, para substituírem as notas fiscais por uma lista das operações realizadas. Na realidade, nos esquemas anteriores, as linhas das listagens eram consideradas e tratadas como notas fiscais o que, em lugar de simplificar sua apuração, tornou-a ainda mais difícil.

Em uma reunião com os responsáveis pelos CPD das firmas, ficou decidido que em substituição às antigas listagens seriam apuradas pelas firmas os resumos de suas operações conforme modelo anexo. O conjunto destes resumos é tratado como estrato sem amostragem que seria simplesmente adicionado aos totais obtidos a partir das notas fiscais em cada Estado.

4.2 - Estratificação

O conjunto das notas fiscais de cada Estado é estratificado, segundo o valor líquido das mercadorias da nota fiscal. A estratificação nos Estados pequenos é de toda população, enquanto nos Estados grandes é a amostra preliminar que é estratificada.

A estratificação se justifica porque o valor das notas fiscais tem distribuição altamente assimétrica, o que ocorre com frequência nos fenômenos econômicos, e nestas condições ela aumenta a eficiência da amostragem.

Examinando a distribuição da população nos estratos definidos para 1974, se conclui que em consequência da inflação não se deve manter fixos os limites dos estratos, sob pena de prejudicar de forma essencial todo processo de amostragem. A fim de superar este inconveniente, adotou-se para limite dos estratos uma série regular, que permitirá a fusão de estratos inferiores e a criação de novos estratos no ponto mais alto da escala sem alteração de sua estrutura básica. Espera-se que desta forma seja possível manter constante a proporção da popula-

ção nos estratos, facilitando desta maneira a comparação ao longo do tempo.

Foi adotada em caráter experimental a série:

10.000 — 30.000 — 100.000 — 300.000 . . .

e posteriormente deverá ser analisada a conveniência de subdividir essas classes, com o objetivo de aumentar a eficiência da amostragem.

O esquema de estratificação está descrito no quadro 1.

Conforme já foi dito anteriormente, as listagens constituem resumo das operações das firmas com contabilidade mecanizada e vão constituir o estrato VI, que como o estrato V tem fração de amostragem igual a 1 (um).

O material tanto dos Estados pequenos como dos grandes é separado segundo os mesmos estratos, divergindo o método de amostragem, porque nos Estados grandes há subamostragem estratificada, enquanto nas Unidades pequenas há simples estratificação sem amostragem preliminar.

A seleção da amostra nos estratos é feita de forma sistemática, com fração variável de amostragem, levando em conta o tamanho pré-determinado da amostra.

QUADRO 1. - Esquema de Estratificação das Notas Fiscais Segundo o Valor Líquido do Total das Mercadorias

Estrato	Intervalo em Cr\$		
I	0	a menos de	10.000
II	10.000	a menos de	30.000
III	30.000	a menos de	100.000
IV	100.000	a menos de	300.000
V	300.000	e	mais
		Listagens	

Os intervalos para a seleção sistemática são calculados levando-se em conta o tamanho N_i do estrato i e o tamanho pré-determinado da amostra n_i . O intervalo é dado pelo inteiro mais próximo de N_i/n_i .

Inicia-se a amostragem com a seleção aleatória da primeira unidade de amostragem no primeiro intervalo, sendo feita a seleção sistemática a partir desta primeira unidade.

4.3 - Nível de Agregação dos Resultados

Está claro que o ideal seria não agregar e produzir estimativas individualizadas por produto e não por grupos de produtos, como os definidos por 3 dígitos da NBM, conforme é proposto pela Fundação IBGE. Entretanto, este ideal só poderia ser atingido através da apuração de todas as notas fiscais recebidas, ou seja, um Censo.

Esta solução é inexecutável tanto pela grande quantidade de meios necessários à apuração mensal de 1,2 milhão de notas fiscais, como pela extensão e complexidade da enorme quantidade de resultados que este esquema proporcionaria.

Torna-se, portanto, necessário encontrar um compromisso entre o ideal das estimativas por produto e a disponibilidade de recursos, tanto para apuração como para publicação dos resultados.

O compromisso consistirá na aplicação do método de amostragem e na elaboração de resultados agregados que façam uso máximo das informações disponíveis. Isto será possível quando a agregação for levada unicamente até o ponto em que a amostra proporcione estimativas dos agregados, com erro de amostragem não superior ao limite considerado aceitável, porque a estimativa de um fenômeno, feita por amostragem só pode ser utilizada com segurança na análise econômica quando o erro a que está sujeita não excede determinado limite.

Parece lógico admitir que uma estimativa do total de uma mercadoria exportada para determinado Estado com erro-padrão superior a 20% não conterà quantidade suficiente de informação para ser aproveitada em uma análise econômica.

Considerando que o mais importante é obter uma boa estimativa anual do comércio por vias internas, decidiu-se entretanto apurar e publicar a cada mês, para cada Estado, o valor e o peso total das mercadorias incluídas em cada uma das oito classes NBM, porque essas estimativas serão produzidas com erro-padrão em percentagem não superior a 10%, quando a amostra for de 50.000 notas fiscais.

O nível mínimo de agregação necessário na elaboração de estimativas anuais só será definido através da análise do conjunto de dados acumulados durante 12 meses e após a adoção de um limite máximo para o erro de amostragem dos totais dos agregados.

Por outro lado, já se dispõe de informação que permite reduzir de 10 para 3 as classes de natureza das operações, e de 7 para 2 as de vias de transporte.

Constatou-se que as vendas e transferências são as operações mais importantes; as demais serão agrupadas em outras operações.

O transporte rodoviário predomina de maneira quase absoluta sobre as outras formas de transporte de mercadorias para os demais Estados. As demais formas de transporte são agrupadas em outras.

Se ao final do ano for possível publicar com segurança informações sobre os 392 agrupamentos de mercadorias definidas pelos 3 dígitos da NBM, haverá para cada Estado 1.960 estimativas de valor e outras tantas de peso, o que constitui, sem dúvida, quantidade apreciável de informações.

Está claro que o nível de agregação para publicação nada tem a ver com o armazenamento das informações, que deve ser feito de forma a registrar toda informação coletada, pois elas constituem informações de valor histórico, que poderão eventualmente ser utilizadas de forma imprevisível.

5 - ESTRUTURA DA AMOSTRA COM SUBAMOSTRAGEM

O grande número de notas fiscais recebidas pela Fundação IBGE é separado em dois grupos, segundo o valor líquido da operação. Atualmente é adotado como ponto de separação entre os grupos o valor de Cr\$100.000,00 (cem mil cruzeiros); em 1974 o limite era Cr\$10.000,00. Este valor deve ser reajustado periodicamente, para que o tamanho da amostra não aumente exageradamente, reduzindo as vantagens da amostragem preliminar sem qualquer compensação. Sendo a inflação da ordem de 40% ao ano, a fim de manter constante a proporção de notas fiscais nos dois grupos, seria necessário transferir o limite de Cr\$100.000,00 para Cr\$140.000,00 ao final de um ano de trabalho.

Do grupo de valor inferior a Cr\$100.000,00 é extraída amostra sistemática de dez por cento.

Na prática, a classificação das notas fiscais nos dois grupos e a amostragem sistemática do grupo de menor valor é feita simultaneamente pela mesma pessoa.

Ao final dessas operações, se obtém a "amostra preliminar", que é constituída dos 10% das notas fiscais do primeiro grupo mais a totalidade das notas fiscais de valor líquido igual ou superior a cem mil cruzeiros. Dessa "amostra preliminar" é que será extraída a subamostra.

Símbolo	Especificação
ℓ	= Via de transporte $\ell = 1, 2, \dots, 7;$
o	= Número de ordem seqüencial no estrato $i, o = 1, 2, \dots, n_i;$
\hat{N}	= Tamanho estimado da população;
N_i	= Tamanho do estrato i na população; \hat{N}_i = sua estimativa;
M	= Tamanho da amostra preliminar;
M_i	= Tamanho da amostra preliminar no estrato $i;$
r_i	= Proporção do estrato i na população; \hat{r}_i = sua estimativa $= \frac{M_i}{M}$
n_i	= Tamanho da subamostra no estrato $i;$
g_i	= Fator de elevação do estrato $i;$
n_{ij}	= Número de vezes que a mercadoria j aparece na amostra do estrato $i;$
n_{ijk}	= Número de vezes que a mercadoria j ocorre no estrato i juntamente com a natureza da operação $k;$
$n_{ij\ell}$	= Número de vezes que a mercadoria j ocorre no estrato i juntamente com a via de transporte $\ell ;$
p_j	= Proporção de registros da mercadoria j no total de registros; \hat{p}_j = Sua estimativa;
p_{ij}	= Proporção de registros da mercadoria j no total de registros do estrato $i;$ \hat{p}_{ij} sua estimativa;
$x_{ijk\ell o}$	= Valor em cruzeiros em uma unidade da amostra;
$y_{ijk\ell o}$	= Peso da mercadoria em uma unidade da amostra;

Símbolo

Especificação

\bar{x}_{ij} = Média do valor da mercadoria j no estrato i da subamostra;

\bar{X}_{ij} = Média do valor da mercadoria j no estrato i da população;

\bar{X}_j = Média do valor da mercadoria j na população;

\bar{Y}_{ij} = Média do peso da mercadoria j no estrato i da população; e

\bar{Y}_j = Média do peso da mercadoria j na população.

6 - ESTIMADORES PARA ESTADOS PEQUENOS (Sem Subamostragem)

A população dos Estados pequenos é recebida em sua totalidade, e a primeira manipulação consiste na estratificação e contagem nos estratos que proporciona os valores N_i = tamanho da população no estrato i.

O tamanho da população no Estado é dado por

$$N = \sum_i N_i$$

Das N_i notas fiscais do estrato i são retiradas sistematicamente n_i que constituem a amostra.

O tamanho da amostra no Estado é dada por

$$n = \sum_i n_i$$

Entre as n_i unidades da amostra do estrato i encontram-se n_{ij} registros referentes à mercadoria j.

A estimativa de p_{ij} , proporção de registros da mercadoria j no total de registros da população de estrato i, é dada por

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i}$$

A estimativa de N_{ij} = número de notas fiscais com a mercadoria j no estrato i da população é dada por

$$\hat{N}_{ij} = \hat{p}_{ij} N_i$$

A variância de \hat{N}_{ij} é estimada por:

$$s^2 \hat{N}_{ij} = N_i (N_i - n_i) \frac{\hat{p}_i (1 - \hat{p}_i)}{n_i - 1}$$

O número total de notas fiscais com a mercadoria j , enviada para o Estado, é estimado por

$$\hat{N}_j = \sum_i \hat{N}_{ij} = \sum_i \hat{p}_{ij} N_i$$

A variância de N_j é estimada por:

$$s^2 \hat{N}_j = \sum_i N_i (N_i - n_i) \frac{1}{n_i - 1} \hat{p}_i \hat{q}_i \text{ onde } \hat{q}_i = 1 - \hat{p}_i$$

6.1 - Estimativa do Valor ou Peso da Mercadoria

Seja $x_{ijk\ell}$ o valor líquido em cruzeiros, no registro $ijk\ell$ o.

O valor total da mercadoria j na amostra do i -ésimo estrato é dado por:

$$t_{ij} = \sum_{o=1}^{n_i} x_{ijk\ell o}$$

Este total permite calcular duas médias distintas a saber: a média aritmética simples

$$\bar{x}_{ij} = \frac{t_{ij}}{n_i}$$

e a média com característica

$$\bar{x}'_{ij} = \frac{t_{ij}}{n_{ij}}$$

As duas médias só coincidem quando $n_i = n_{ij}$, isto é, quando todos os registros do estrato i se referem à mercadoria j .

A estimativa do total do estrato T_{ij} é dada por:

$$\hat{T}_{ij} = N_i \bar{x}_{ij} = \frac{N_i \sum_{k=0}^{n_i} x_{ijk} \ell_o}{n_i}$$

A variância deste total é estimada por:

$$s^2_{\hat{T}_{ij}} = N_i (N_i - n_i) \frac{1}{n_i} s^2_{x_{ij}} = N_i (N_i - n_i) \frac{\sum_{k=0}^{n_i} x^2_{ijk} \ell_o - \frac{t^2_{ij}}{n_i}}{n_i (n_i - 1)}$$

A estimativa do total da mercadoria j no Estado é dada por:

$$\hat{T}_j = \sum_i \hat{T}_{ij} = \sum_i N_i \bar{x}_{ij}$$

A variância deste total é dada por:

$$s^2_{\hat{T}_j} = \sum_i s^2_{\hat{T}_{ij}}$$

As fórmulas de totais e variâncias para subclasses referentes à natureza da operação ou via de transporte são equivalentes, com as

mudanças apropriadas nos subíndices.

7 - ESTIMADORES PARA ESTADOS GRANDES (Com Subamostragem)

Na amostragem dupla, a proporção de elementos da população em cada estrato é desconhecida, e sua estimativa é feita a partir de uma amostra preliminar retirada ao acaso ou de forma sistemática. As características de interesse são observadas em uma subamostra selecionada da amostra preliminar.

Seja M_i o tamanho do estrato i na amostra preliminar. O tamanho do estrato correspondente na população N_i é desconhecido, devendo portanto ser estimado.

No esquema de amostragem adotado para os Estados grandes, a amostra preliminar constituiu-se de 10% das notas fiscais de valor líquido inferior a Cr\$100.000,00, selecionadas de forma sistemática.

Das notas fiscais de valor líquido igual ou superior a Cr\$100.000,00 não se retira amostra preliminar. O tamanho da amostra preliminar e a estimativa do tamanho da população nos estratos estão esquematizados no quadro 2.

QUADRO 2. - Tamanho da Amostra Preliminar e Estimativa de Tamanho da População nos Estados

Nº	Estrato		Tamanho da amostra preliminar	Estimativa do tamanho da população
	Intervalo (Cr\$1.000)			
I	0	10	M_1	$\hat{N}_1 = 10 M_1$
II	10	30	M_2	$\hat{N}_2 = 10 M_2$
III	30	100	M_3	$\hat{N}_3 = 10 M_3$
IV	100	300	M_4	$\hat{N}_4 = M_4$
V	300 e mais		M_5	$\hat{N}_5 = M_5$
VI	Listagens		M_6	$\hat{N}_6 = M_6$

Portanto, a estimativa de N , número total de notas fiscais do Estado, é feita por

$$\hat{N} = \sum_i \hat{N}_i$$

A subamostra n_i é retirada do i -ésimo estrato da amostra preliminar.

Entre as n_i unidades de amostragem do i -ésimo estrato encontram-se n_{ij} registros referentes à mercadoria j .

A estimativa de p_{ij} , proporção de registros da mercadoria j no total de registros do estrato i , é dada por

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i}$$

A estimativa de N_{ij} , número de notas fiscais com a mercadoria j no estrato i da população é dada por:

$$\hat{N}_{ij} = \hat{p}_{ij} \hat{N}_i$$

De maneira geral o tamanho da amostra preliminar é dado por

$$M = \sum_i M_i \text{ e a razão } \hat{r}_i = \frac{M_i}{M} \text{ constitui estimativa não viesada de } r_i = \frac{N_i}{N}$$

proporção dos elementos do estrato i na população.

O número total de operações de exportação da mercadoria j para o Estado é estimado por

$$\hat{N}_j = \sum_i \hat{N}_{ij} = \sum_i \hat{p}_{ij} \hat{N}_i$$

7.1 - Estimativa do Valor e do Peso Total de uma Mercadoria

Sejam x_{ijkl} o e y_{ijkl} o valor líquido em cruzeiros e o peso líquido, respectivamente, no registro $i j k l$ o.

O valor total da mercadoria j na subamostra do i -ésimo estrato é dado por

$$t_{ij} = \sum_0 x_{ij} \dots 0$$

onde as n_i unidades da subamostra são escrutinadas somando o valor quando a mercadoria é j , independentemente da natureza da operação e da via de transporte.

Este total permite calcular duas médias distintas, a saber: a média aritmética simples e a média com a característica, isto é, o valor médio dos registros em que aparece a mercadoria j ;

$$\bar{x}_{ij} = \frac{t_{ij}}{n_i} \quad \text{e} \quad \bar{x}'_{ij} = \frac{t_{ij}}{n_{ij}}$$

A média ponderada \bar{x}_j obtida através da subamostra é dada por:

$$\bar{x}_j = \sum_i \hat{r}_i \bar{x}_{ij}$$

e proporciona estimativa não viesada da média \bar{X}_j da população, pois:

$$\begin{aligned} E(\bar{x}_j) &= \sum_i E(\hat{r}_i \bar{x}_{ij}) = \sum_i E(\hat{r}_i E[\bar{x}_{ij} | \hat{r}_i]) \\ &= \sum_i E(\hat{r}_i \bar{X}_{ij}) = \sum_i r_i \bar{X}_{ij} = \bar{X}_j \end{aligned}$$

A estimativa do total do estrato T_{ij} pode ser feita através de qualquer uma das médias e o multiplicador correspondente, e é dada por

$$\hat{T}_{ij} = \hat{N}_i \bar{x}_{ij} = \hat{N}'_{ij} \bar{x}'_{ij} = \frac{\sum_0 x_{ij} \dots 0}{n_i} \hat{N}_i$$

A variância da média ponderada é dada por:

$$\begin{aligned} V(\bar{x}_j) &= E \left[V \left(\sum_i \hat{r}_i \bar{x}_{ij} \mid \hat{r}_i \right) \right] + V \left[E \left(\sum_i \hat{r}_i \bar{x}_{ij} \mid \hat{r}_i \right) \right] \\ &= E \left[\sum_i \hat{r}_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i} \right) S_{ij}^2 \right] + V \left[\sum_i \bar{X}_{ij} \hat{r}_i \right] \end{aligned}$$

Sendo S_{ij} e \bar{X}_{ij} parâmetros, têm-se:

$$V(\bar{x}_j) = \sum_i E(\hat{r}_i^2) \left(\frac{1}{n_i} - \frac{1}{N_i} \right) S_{ij}^2 + \sum_i \bar{X}_{ij}^2 V(\hat{r}_i) + \sum_{i \neq h=1} \bar{X}_{ij} \bar{X}_{hj} \text{Cov}(\hat{r}_i, \hat{r}_h)$$

$$\begin{aligned} V(\bar{x}_j) &= \sum_i \left(\frac{1}{n_i} - \frac{1}{N_i} \right) \hat{r}_i^2 S_{ij}^2 + \sum_i \left[\left(\frac{1}{n_i} - \frac{1}{N_i} \right) S_{ij}^2 + \bar{X}_{ij}^2 \right] V(\hat{r}_i) + \\ &\quad + \sum_{i \neq h=1} \bar{X}_{ij} \bar{X}_{hj} \text{Cov}(\hat{r}_i, \hat{r}_h) \end{aligned}$$

A variância e a co-variância das proporções \hat{r}_i e $\hat{r}_i \hat{r}_h$ são dadas por:

$$V(\hat{r}_i) = \frac{N-M}{N-1} \frac{\hat{r}_i(1-\hat{r}_i)}{M}; e$$

$$\text{Cov}(\hat{r}_i, \hat{r}_h) = \frac{M-N}{N-1} \frac{\hat{r}_i \hat{r}_h}{M}$$

Substituindo na fórmula da variância da média e simplificando, tem-se:

$$V(\bar{x}_j) = \sum_i \left(\frac{1}{n_i} - \frac{1}{N_i} \right) \hat{f}_i^2 S_{ij}^2 + \frac{M-N}{N-1} \frac{1}{M} \left[\sum_i \left(\frac{1}{n_i} - \frac{1}{N_i} \right) \hat{f}_i (1 - \hat{f}_i) S_{ij}^2 + \sum_i \hat{f}_i (\bar{X}_{ij} - \bar{X}_j)^2 \right]$$

No segundo membro tem-se duas parcelas; a primeira é a variância da média de uma amostra estratificada, quando se conhece o tamanho dos estratos, e a segunda estima o aumento da variância provocado pela necessidade de se estimar o tamanho dos estratos, através da amostra preliminar.

Se na segunda parcela for desprezada a correção para a população finita e o somatório da variância nos estratos, que é pequena em relação ao termo seguinte, será obtida uma estimativa aproximada do aumento de variância, provocado pelo desconhecimento do tamanho dos estratos na população. A estimativa do aumento da variância é dada por:

$$\frac{1}{M} \sum_i \hat{f}_i (\bar{X}_{ij} - \bar{X}_j)^2$$

Com esta simplificação, a estimativa da variância da média ponderada será dada por:

$$V(\bar{x}_j) \approx \sum_i \frac{\hat{f}_i^2 S_{ij}^2}{n_i} + \frac{1}{M} \sum_i \hat{f}_i (\bar{X}_{ij} - \bar{X}_j)^2$$

8 - OTIMIZAÇÃO DA AMOSTRA

Foi feita até aqui uma exposição sobre os problemas de definição da população e das unidades de amostragem, bem como descrição da estrutura das amostras derivadas com o auxílio das informações fornecidas pelos levantamentos do início de 1974, que foram utilizados como "levantamento piloto" para esta pesquisa. A seguir,

foram dadas as estimativas para cada estrutura.

Isto era o máximo possível com as informações disponíveis no início e aquelas acumuladas ao longo da pesquisa.

Entretanto, em cada levantamento mensal, informações novas e cada vez mais precisas sobre a estrutura da população e os componentes da variância e dos custos poderão ser utilizadas no aperfeiçoamento do esquema de amostragem.

O ideal será que juntamente com os resultados de cada levantamento seja redimensionada a amostra do levantamento seguinte.

A meta a atingir será a obtenção do máximo de desagregação, no que se refere à informação anual sobre mercadorias, levando em conta as restrições sobre custo.

8.1 - Condição para Adoção da Subamostragem

A seguir, é tratado o problema de otimização de amostra com amostragem dupla.

Como foi visto, uma estimativa aproximada da variância da média ponderada é dada por:

$$V(\bar{x}_j) \cong \sum_i \frac{f_i^2 S_{ij}^2}{n_i} + \frac{1}{M} \sum_i f_i (\bar{X}_{ij} - \bar{X}_j)^2$$

Por outro lado, a variância da média da amostra quando a amostragem é irrestrita aleatória, isto é, sem estratificação, é dada por:

$$V(\bar{x}_j) = \frac{N-n}{N} \frac{S^2}{n} \cong \frac{\sum_i n_i S_i^2 + \sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2}{n}$$

Comparando esta variância com a variância da média ponderada na subamostragem, verifica-se que a diferença entre ambas pode ser positiva, nula ou negativa, dependendo da alocação da subamostra nos estratos.

Entretanto, o problema importante na amostragem dupla consiste em saber se o custo da amostra preliminar compensa a redução da variância, não importando a comparação da variância de suas estimativas com as variâncias das estimativas correspondentes da amostragem irrestrita aleatória.

Colocando de outra forma, pergunta-se: a um custo fixo é mais conveniente extrair uma amostra irrestrita aleatória ou uma amostra dupla?

Para responder a esta questão, considerou-se uma função de custo simples:

$$C = c_1 M + c_2 \sum_i n_i,$$

onde:

c_1 = custo por unidade da amostra preliminar;

M = tamanho da amostra preliminar;

c_2 = custo de uma unidade na subamostra; e

n_i = tamanho da subamostra no estrato i .

Para uma amostra irrestrita aleatória, o plano ótimo é dado por:

$$n' = \frac{\text{Custo total préfixado}}{\text{Custo por unidade de amostragem}} = \frac{C_0}{c_2}$$

Na amostragem dupla, a estimativa da variância da média depende do tamanho relativo dos estratos na população (r_i), das médias e variâncias nos estratos (\bar{x}_{ij} , s_{ij}^2), e também do tamanho da amostra preliminar (M) e da alocação da subamostra nos estratos (n_i).

Os valores ótimos de M e n_i são aqueles para os quais a variância é mínima sujeita à condição:

$$C = c_1 M + c_2 \sum_i n_i.$$

A técnica para a determinação de M e dos n_i ótimos consiste em resolver a função:

$$F = \sum_i \frac{r_i^2 S_i^2}{n_i} + \frac{1}{M} \sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2 + \mu (c_1 M + c_2 \sum_i \mu_i)$$

$$= \sum_i \left(\frac{r_i S_i}{n_i} \sqrt{\mu c_2 n_i} \right)^2 + \left(\frac{1}{M} \sqrt{\sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} - \sqrt{\mu c_1 M} \right)^2 + I,$$

onde: I = termos independentes de M e n_i , assim:

$$I = 2 \sqrt{\mu c_2} \sum_i r_i S_i + 2 \sqrt{\mu c_1} \sqrt{\sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2}$$

As soluções desta equação são:

$$\frac{r_i S_i}{\sqrt{n_i}} = \sqrt{\mu c_2 n_i} \quad \therefore \quad n_i = \frac{r_i S_i}{\sqrt{\mu c_2}} \quad ; e$$

$$\sqrt{\frac{\sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2}{M}} = \sqrt{\mu c_1 M} \quad \therefore \quad M = \sqrt{\frac{\sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2}{\mu c_1}}$$

O valor de μ é determinado quando a variância é minimizada para um custo pré-fixado ou então quando o custo é minimizado para uma variância pré-fixada.

Para que as estimativas possam ser utilizadas com segurança nas análises econômicas, é indispensável que seus erros de amostragem não excedam limites pré-fixados. Em consequência, torna-se necessário minimizar o custo com variância pré-fixada, substituindo as expressões ótimas de M e n_i na fórmula simplificada da variância da média:

$$V(\bar{x}_j) \cong \sum_i \frac{r_i^2 S_i^2}{n_i} + \frac{1}{M} \sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2 =$$

$$= \sum_i r_i^2 S_i^2 \frac{\sqrt{\mu} c_2}{r_i S_i} + \frac{\sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2}{\sqrt{\frac{\sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2}{\mu c_1}}}$$

$$= \sqrt{\mu c_2} \sum_i r_i S_i + \sqrt{\mu c_1} \sqrt{\sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} \quad \therefore$$

$$\sqrt{\mu} = \frac{V_0}{\sqrt{c_2} \sum_i r_i S_i + \sqrt{c_1} \sqrt{\sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2}}$$

onde V_0 = variância pré fixada.

Substituindo o valor de μ nas expressões ótimas de M e n_i tem-se:

$$n_i = \frac{r_i S_i}{\sqrt{c_2}} \times \frac{1}{V_0} \left[\sqrt{c_2} \sum_i r_i S_i + \sqrt{c_1} \sqrt{\sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} \right] ; e$$

$$M = \frac{1}{V_0 \sqrt{c_1}} \left[\sqrt{\sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} \right] \left[\sqrt{c_2 \sum_i r_i S_i} + \sqrt{c_1 \sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} \right]$$

Substituindo M e n_i na expressão de custo da amostragem dupla:

$$C_d = c_1 M + c_2 \sum_i n_i ,$$

tem-se:

$$C_d = c_1 \frac{1}{V_0 \sqrt{c_1}} \left[\sqrt{\sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} \right] \left[\sqrt{c_2 \sum_i r_i S_i} + \sqrt{c_1 \sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} \right] \\ + c_2 \sum_i \frac{r_i S_i}{\sqrt{c_2}} \times \frac{1}{V_0} \left[\sqrt{c_2 \sum_i r_i S_i} + \sqrt{c_1 \sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} \right]$$

Simplificando tem-se:

$$C_d = \frac{1}{V_0} \left[\sqrt{c_2 \sum_i r_i S_i} + \sqrt{c_1 \sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} \right]^2$$

Por outro lado, desprezando o fator de correção para população finita, a variância da estimativa da média na amostragem irrestrita aleatória é dada por:

$$V(\bar{x}_j) \cong \frac{S^2}{n'} \quad \therefore \quad C = c_2 \frac{S^2}{V_0}$$

A amostragem dupla será preferível quando o custo para estimar \bar{X}_j com variância V_0 for inferior ao custo de uma amostra irrestrita aleatória equivalente, ou seja, quando for satisfeita a desigualdade:

$$\frac{1}{V_0} \left[\sqrt{c_2 \sum_i r_i S_i} + \sqrt{c_1 \sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} \right]^2 < \frac{1}{V_0} c_2 S^2 \quad \therefore$$

$$\left[\sqrt{c_2 \sum_i r_i S_i} + \sqrt{c_1 \sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} \right]^2 - (S\sqrt{c_2})^2 < 0 \quad \therefore$$

$$\begin{aligned} & (\sqrt{c_2 \sum_i r_i S_i} + \sqrt{c_1 \sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} + S\sqrt{c_2}) (\sqrt{c_2 \sum_i r_i S_i} + \\ & + \sqrt{c_1 \sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} - S\sqrt{c_2}) < 0. \end{aligned}$$

Para que isto ocorra, o segundo termo do primeiro membro tem que ser menor que zero, então:

$$\sqrt{c_2 \sum_i r_i S_i} + \sqrt{c_1 \sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} - S\sqrt{c_2} < 0$$

$$\sqrt{c_1 \sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} - \sqrt{c_2} (S - \sum_i r_i S_i) < 0$$

$$\sqrt{c_1 \sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2} < \sqrt{c_2} (S - \sum_i r_i S_i)$$

$$c_1 < c_2 \frac{(S - \sum_i r_i S_i)^2}{\sum_i r_i (\bar{X}_{ij} - \bar{X}_j)^2}$$

que é condição necessária para que compense a retirada da amostra preliminar, ou seja, a adoção da amostragem dupla.

8.2 - Construção dos Estratos

Considerando o processo inflacionário a que está submetida nossa economia, será necessário reavaliar anualmente a eficiência da estratificação que está sendo adotada, a fim de decidir sobre a conveniência de mudar os limites dos estratos.

Não se dispõe, infelizmente, de uma técnica válida para todos os casos. Para nosso esquema de estimação, em que a média é ponderada em função do tamanho dos estratos, há algumas alternativas possíveis, entre as quais COCHRAN (1) selecionou duas técnicas que se mostraram adequadas em todas as aplicações feitas por ele.

A primeira técnica foi desenvolvida por Dalenius e Hodges. Nela são calculados os valores acumulados de $\sqrt{f(y)}$. Chamando de R a soma daqueles valores em todo intervalo de variação da variável, então definiu-se os limites dos estratos por:

$$Y_i = \frac{iR}{L} \quad | \quad i = 1, 2, \dots, (L-1) \quad |$$

Na técnica de Ekman, desenvolvida posteriormente, os limites dos estratos são definidos de tal modo que satisfaçam a seguinte equação:

$$\frac{M_i}{M} (Y_i - Y_{i-1}) = \text{constante}$$

onde M é o tamanho da amostra preliminar.

Qualquer dessas duas técnicas define um conjunto de limites dos estratos próximos dos conjuntos de pontos de estratificação ótima.

8.3 - Aplicação da Programação não Linear na Alocação da Amostra nos Estratos

Os estimadores apresentados anteriormente permitem determinar o tamanho da amostra e sua alocação nos estratos (cálculo dos n_i), uma vez fixado o erro de amostragem aceitável para a estimativa do total de determinada mercadoria e o nível de confiança correspondente.

Entretanto, como foi visto, em cada Estado podem ocorrer até 1091 mercadorias e, evidentemente, o comportamento da variância nos estratos pode ser muito diferente para diferentes mercadorias. Nestas condições, uma alocação da amostra aceitável para um grupo de mercadoria poderá comprometer a exatidão dos resultados de outro grupo. Este fato torna indispensável um compromisso entre as variâncias para determinar uma alocação "ótima" da amostra.

JAGANHATHAN (5) cita a proposta de Seth com aplicação da programação não linear para resolver o problema geral de alocação da amostra.

Chame-se:

$i = 1, 2, \dots, 5, 6$ os estratos de tamanho; e

$j = 1, 2, \dots, 1091$ as mercadorias segundo a TAB.

Como foi visto, a variância da estimativa do total da mercadoria j é dada por:

$$s_{T_j}^2 = \sum_i N_i (N_i - n_i) \frac{s_{ij}^2}{n_i} + \frac{1}{M^2} \sum_i M_i (\bar{X}_{ij} - \bar{x}_j)^2$$

Para uma função de custo:

$$C = M c_1 + \sum_i n_i c_2 ,$$

em que:

M = tamanho da amostra preliminar;

c_1 = custo de uma unidade da amostra preliminar;

c_2 = custo de uma unidade da subamostra, considerando que o custo não varia entre os estratos; e

n_i = tamanho da subamostra no estrato i .

O custo deverá ser minimizado sujeito às seguintes restrições:

$$n_i \leq M_i ; e$$

$$s_j \leq d_j .$$

Os d_j são fixados em função da exatidão que se deseja para a estimativa do total da mercadoria j e o nível de confiança correspondente.

Seth propõe a seguinte transformação:

$$l_i = \frac{1}{n_i} - \frac{1}{M_i} ,$$

que permite a solução do problema pelo método de programação não linear.

O problema se apresenta então da seguinte forma:

$$\text{minimizar } F(l_1, l_2, \dots, l_5, l_6) = \sum_i \frac{c_2}{l_i + h_i} \quad (h_i = \frac{c_2}{M_i}),$$

sujeita às seguintes restrições:

$$\sum_i s_{ij}^2 \times l_i \leq p_j, \text{ em que:}$$

$$l_i \geq 0 ; e \text{ onde}$$

$$p_j = d_j - \left[\frac{1}{M^2} \sum_i M_i (X_{ij} - X_j)^2 + \sum_i h_i s_{ij}^2 \right] ,$$

é a variância máxima admissível, considerando o erro de amostragem e o nível de confiança escolhidos.

LITERATURA CONSULTADA

- 1 - COCHRAN, W. G. *Sampling techniques*. New York, J. Wiley. 1960. 330 p.
- 2 - DEMING, W. E. *Some theory of sampling*. New York, J. Wiley. 1950. 602p.
- 3 - HEADY, E. O. & CANDLER, W. *Linear programming methods*. Ames, Iowa State University, 1958. 597p.
- 4 - JAGANNATHAN, R. A method for solving a nonlinear programming problem in sample surveys. *Econometrica*, Clevedon, 33 (4): 841-846, Oct. 1965.
- 5 - _____ . The programming approach in multiple character studies. *Econometrica*, Clevedon, 33 (1): 236-237, Jan. 1965.
- 6 - JENNERGREN, P. A price schedule decomposition algorithm for linear programming problems. *Econometrica*, Clevedon, 41 (5): 965-980, sept. 1973.
- 7 - KISH, Leslie. *Survey sampling*. New York, J. Wiley, 1950. 602p.
- 8 - MADOW, W. G. *Teoria dos levantamentos por amostragem*. Rio de Janeiro, IBGE, 1951. 256p.
- 9 - SUKHATAME, B. V. & P. V. *Sampling theory of surveys with applications*. 2nd. ed. Roma, Nações Unidas, FAO, 1970. 452p.
- 10 - WEISS, L. *Statistical decision theory*. New York, McGraw-Hill, 1961. 195p.
- 11 - YATES, Frank. *Sampling methods for censures and surveys*. London, Charles Griffin, 1953. 401p.

RESEARCH OF A SAMPLING METHOD TO ESTIMATE INDIVIDUAL ITEMS VALUE OF BRAZILIAN STATES COMMERCE

SUMMARY

This study describes a research done at the Divisão de Estatísticas Econômicas - Coordenadoria de Análise de Dados - Secretaria de Economia e Planejamento, started in 1976. The objective was to find an operational scheme which would allow the estimation of commodity exports from the State of São Paulo to other Brazilian States.

These estimates are very important for regional economic analysis and planning. Given the large number of commercial operations that constitute this flow of goods from the State of São Paulo - more than one million per month - it was necessary to use sampling to obtain the needed estimates. Hence, the first step was to investigate the best sampling technique. Thus, the transactions with each one of the other states were taken as independent populations and the fiscal note was taken as sampling unit.

Practical and theoretical reasons indicated that stratified sampling was the best technique to be used for 13 States where the number of fiscal notes is small, while double stratified sampling was better for the remaining 12 States.

A pilot survey was done using the 1974 survey data in order to estimate the starting sample size. The sample size, necessary to estimate the total of 8 commodity categories from the Nomenclatura Brasileira de Mercadorias for each month and State, with a standard error equal to 10% or less was 50.000 fiscal notes.

The procedure and the practical problems encountered in manipulating that large sample are described. Later on, two sampling schemes and the problems which arise when they are applied, are described. A detailed analysis of the appropriate estimators is also presented.

This study also presents the statistical methods used to optimize the State samples. This optimization relates to the convenience to adopt subsampling; to estimate the strata intervals; and finally to establish the compromise of the variances of the several commodities since this is a multiple characteristic sampling procedure.

Given the dynamic character of the States economy, it is expected that its commerce with the other States will change with respect to both, type and quantity of commodities exported. This indicates the need for a sampling method flexible enough to adjust itself to the new changes, when the sample efficiency decreases below acceptable levels.