

FUNDAÇÃO IBGE
INSTITUTO BRASILEIRO DE ESTATÍSTICA

COMISSÃO "B"
- Contribuição -
nº 1

AMOSTRAGEM APLICADA À ESTATÍSTICA DO COMÉRCIO
POR VIAS INTERNAS

Autor: Contribuição do Departamento Estadual de Estatística do Estado de São Paulo.



PRIMEIRA CONFERÊNCIA NACIONAL DE ESTATÍSTICA

29 de maio a 4 de junho de 1968

Rio de Janeiro — GB

AMOSTRAGEM APLICADA À ESTATÍSTICA DO COMÉRCIO POR VIAS INTERNAS

Contribuição do Departamento de Estatística do Estado de São Paulo à Primeira Conferência Nacional de Estatística.

A estatística do comércio entre as unidades federadas, por vias internas, constituiu sempre problema de difícil solução. No caso específico de São Paulo, dado o grande volume e a variedade de mercadorias vendidas às outras Unidades da Federação, a matéria assume significação ímpar.

Rio de Janeiro, Guanabara
1968

O dinamismo da economia paulista e o fato de haver ao lado da agricultura uma indústria diversificada fizeram com que o Estado de São Paulo se tornasse o principal fornecedor de alimentos e produtos manufaturados para todo o Brasil. Considere-se ainda que a localização do Estado, com fronteiras abertas ao sul, ao leste, ao centro-oeste e ao oeste, facilita o intercâmbio comercial, por haver ligações rodoviárias e ferroviárias em boas condições técnicas.

O elevado número de centros urbanos no interior de São Paulo, alguns com grande presença das funções industrial e comercial, criou condições para ampliar o comércio com outros Estados. Podemos notar a presença de alguns polos de onde partem as exportações para outras Unidades, principalmente devido à localização estratégica de tais centros em relação às direções rodoferroviárias.

A região que tem como polo Ribeirão Preto, por exemplo, mantém ativo comércio com o Triângulo Mineiro e Sul de Goiás. A região da Alta Sorocabana, tendo como centro Presidente Prudente, mantém intercâmbio com o Norte do Paraná e Sul de Mato Grosso. Os municípios do Vale do Paraíba comerciam intensamente com os Estados do Rio de Janeiro e Guanabara.

Levando em conta essas observações, foi prevista a possibilidade de se conhecer também os fluxos de mercadorias segundo a região de procedência.

A importância do transporte rodoviário na atualidade é bem conhecida, dispensando comentários sobre a sua significação no incremento das transações comerciais.

O grande problema a equacionar na estatística do comércio por vias internas era o da fonte de dados, que necessitava ser legal, permanente e segura. Os Decretos ns. 60.467, 60.887 e 62.158, respectivamente de 14.3.1967, 22.6.1967 e 19.1.1968, abriram oportunidade para o trabalho.

Essa nova legislação, se de um lado abriu perspectivas para a implantação do levantamento, por outro colocou o Departamento de Estatística em situação difícil, pois o volume previsto

de notas fiscais, emitidas cada mês, atingiria três milhões. De fato, o trabalho de controle de recebimento, crítica, codificação e perfuração de cartões referentes a tal volume de notas requer elevado número de funcionários, sendo de custo operacional tão alto, que torna a tarefa impraticável.

O Departamento de Estatística, há algum tempo, tentando evitar o tratamento extensivo e na ausência de outra fonte de dados, experimentou construir uma amostra com a utilização dos manifestos das empresas de transporte. Nos manifestos são registradas as mercadorias exportadas, especificadamente, uma em cada linha. Podemos afirmar que a experiência foi bem sucedida, apesar de dificuldades que, pela exiguidade do tempo, ainda não puderam ser superadas.

A utilização da nota fiscal impede-nos, agora, de prosseguir na experiência, por razões técnicas e também pela dificuldade de manter a coleta de manifestos com regularidade, uma vez que as empresas têm conhecimento da disponibilidade da nota fiscal para fins estatísticos.

Face ao exposto, decidiu o Departamento de Estatística utilizar uma amostra das notas fiscais, como o recurso mais prático no tratamento de material tão volumoso. A construção de amostra representativa é empreendimento que apresenta alguns obstáculos de difícil superação. Considerando a magnitude do problema, recorreremos a consultores de alto nível, que se prontificaram a participar da elaboração do plano. Assim, estão colaborando com o Departamento de Estatística o professor Lindo Fava, da cadeira de Estatística da Faculdade de Filosofia da Universidade de São Paulo e representante dessa Universidade na Junta Executiva Regional de Estatística, e o dr. Moysés Kessel, consultor estatístico de grande empresa privada e destacado técnico que já integrou os quadros do Instituto Brasileiro de Geografia e Estatística.

O Dr. Moysés Kessel elaborou o plano que será adotado no tratamento das notas fiscais referentes às operações comerciais iniciadas a 1º de janeiro de 1968. A metodologia está exposta a seguir.

Cumpra lembrar que o universo de notas é ainda desconhecido e que não se dispõe do cadastro dos estabelecimentos comerciais emitentes das notas - dois grandes óbices para a construção de uma boa amostra.

METODOLOGIA

Em face do elevado número de guias e da natureza do documento a ser utilizado (nota fiscal), que registra pêsos e valores das mais diversas mercadorias, a primeira preocupação foi dividir em dois o universo global ainda desconhecido, de maneira a aumentar a representatividade dos estratos.

Por conveniência administrativa, visto ser a seleção descentralizada, julgou-se útil reduzir a dois o número de estratos.

O corte no universo já implica numa grande redução da variância, que não se comporta igualmente nos dois subuniversos.

Como as notas fiscais são numeradas, pode-se adotar (experiência da Guanabara e Nordeste), para a escolha das notas o número final da guia, tendo-se assim uma seleção sistemática - o que significa uma vantagem.

Teoricamente, estratos que representam notas de pequeno valor deveriam ter fração de amostragem muito pequena. Ocorre todavia, na prática, que a estratificação por valor, no caso em apreço, implica uma estratificação por mercadoria.

Desta forma, na prática, a fração menor da amostragem deverá ser um pouco maior do que determinada teoricamente, porque a excessiva diversidade de produtos de pequeno valor poderá elevar a variância neste estrato, convindo lembrar que ainda não se tem indício da tendenciosidade.

De acôrdo com uma lei empírica, enunciada por Hansen, Madow e Hurwitz, podemos considerar como extrato superior (100%) aquêle que representa 50% do valor total.

Numa amostra por nós estudada, com base em linhas de manifestos de emprêsas de transporte, verificamos que notas fis-

cais com valor acima de NCr\$ 5 000 atingiram 43% do universo. Baixando o limite de estratificação de NCr\$ 5 000 para NCr\$ 4 000 encontramos 55% do universo, o que atende às condições sugeridas pela lei citada.

A tabela abaixo apresenta a distribuição de frequência do número e dos valores obtidos segundo a tabulação dos dados extraídos dos referidos manifestos e das previsões face ao suposto universo.

Classes de valor NCr\$	Ni	Valor real	% valor	% nº	Nº previsto de notas da amostra preliminar	amostra final objetivo
					10%	5%
0-4000	1240	818 136	44,5	94,7	284 100	142 050
+ 4000	69 1	021 692	55,5	5,3	159 000	159 000
Total	13091	839 828	100	100	443 100	201 100

Suposto universo: 3 000 000 de notas mensais

Para o planejamento do esquema preliminar, o conhecimento de algumas características da distribuição de notas fiscais permitiu estabelecer o tamanho da amostra inicial. Uma forma sumária de obter o tamanho aproximado, calculado pela alocação ótima, consiste em alocar a cada estrato uma proporção da amostra correspondente à proporção do total estimado dentro desse estrato. Em nosso caso, isto implicaria em 55% das unidades da amostra no estrato I (mais de NCr\$4 000) e 45% no estrato II. Como esta regra tende a sub-representar o estrato de notas de menor valor, costuma-se, arbitrariamente, aumentar a representação do estrato menor por um fator em volta de 2. Assim ou faria a amostra total corresponder a cerca de 14% do universo, aproximando-se da fração obtida, após cálculos demorados dos parâmetros dos estratos, no esquema do Prof. Toledo Piza elaborado em 1957.

Se o objetivo do levantamento fôsse sòmente a apuração do valor global da exportação do Estado por vias internas, o planejamento da amostra seria relativamente simples. Ocorre que interessa obter estimativas para grande número de subpopulações: Unidades da Federação, classes de mercadoria e mercadorias principais, em tonelagem e valor.

Algumas dessas subpopulações correspondem a parcela muito pequena do universo das exportações; os Estados e Territórios das regiões Norte e Meio-Norte e alguns do Nordeste representam, individualmente, menos de 1% da quantidade e valôres exportados. A obtenção de estimativas fidedignas implica em aumentar consideravelmente o tamanho da amostra. Outra solução consiste em separar e apurar totalmente as notas com destino a essas Unidades, aplicando a amostragem nos Estados restantes; isto é feito na Guanabara e o foi também no esquema Piza.

Por outro lado, o valor da nota é utilizado como variável de estratificação, mas necessitamos também de estimativas de volume físico, muito importantes para o planejamento da política de transportes.

A distribuição de pêsos das notas é correlacionada com a de valor, dentro de cada classe de mercadorias, mas a distribuição total é menos assimétrica devido ao limite de carga dos caminhões, principal meio de transporte. Na prática, isto implica em elevar a variância do pêsos, devido a inclusão de notas de valor menor mas com grande tonelagem (carga a granel, produtos agrícolas, por ex.), no estrato II.

Aqui também poder-se-ia separar as notas de grande pêsos, após a seleção das de grande valor. Assim se faz na Guanabara. No esquema Piza não se apurou o pêsos das exportações.

Como já foi explicado, qualquer separação adicional de milhões de notas implica em grandes recursos de mão-de-obra e tempo. Tentamos examinar, com êste esquema inicial, a possibilidade de obter dados precisos, com a simples separação de notas de grande valor e a amostragem das restantes. Isto porque a precisão de uma amostra depende muito mais do número de unidades incluídas (ta

manho absoluto) do que da proporção do universo relacionado. Este fato, nem sempre compreendido por leigos, faz com que, em condições similares, uma amostra de 100.000 notas do universo de 1.000.000 (10%) seja muito mais precisa do que uma amostra de 10.000 sôbre 100.000 (também 10%).

O grande número de notas, na amostra em estudo, destina-se exatamente a permitir estimativas precisas de subpopulações; quanto ao volume físico, a separação de notas de grande valor já implica em reduzir bastante a variância das notas sujeitas a amostragem, embora em escala menor do que prevê o valor.

As distribuições para os Estados do Norte-Nordeste são aproximadas às do valor total, com forte assimetria. A separação de notas de maior valor (há poucas notas de grande pêsso e pouco valor para o Norte, devido à importância do frete) implica em reduzir a variância ao ponto de permitir, a grosso modo, estimar com precisão aceitável uma característica que apareça em cerca de 500 notas, ou cerca de 2% do estrato II. Talvez seja necessário grupar, nas tabulações, alguns Estados e publicar resultados englobados para a Região Norte e Meio-Norte.

A confirmação das hipóteses de trabalho do esquema inicial dependerá evidentemente de análise das distribuições de amostragem obtidas, já em andamento.

Esta exposição procurou mostrar a dificuldade de aplicação de modelos matemáticos rígidos a uma pesquisa real. A amostragem, além de ciência, é arte. O eminente estatístico norte-americano Thomaz Nosimarm, muito conhecido dos brasileiros que se aperfeiçoaram nos Estados Unidos, resume assim esta questão:

"A estatística matemática enuncia propriedades de entidades matemáticas e tem exatidão constante e imutável. Em contraste, a estatística real é tanto arte como ciência. Ela cuida de medidas e contagens de objetos físicos e difere da primeira em dois aspectos. Primeiro, Estatística Real é arte, pois medir e contar de maneira científica é uma arte. O estatístico, na vida real, tem que saber como os dados que utiliza foram obtidos; frequentemente êle próprio os obtém. Segundo, como a Estatísti-

ca Real é parte da Matemática Aplicada, êle faz essencialmente e necessariamente algo que a Matemática Pura e a Estatística Matemática nunca fazem - afirmações, em termos quantitativos, sôbre objetos específicos" (Estatística, Junho 1957 - IASI).

A amostra ora construída segundo a exposição que acaba de ser feita possibilitará, gradativamente, o fornecimento de dados sôbre o volume e o valor da exportação, por destino e segundo classes e subclasses de mercadorias.

É evidente que, à medida que forem sendo conhecidos os erros prováveis, cada mês poderemos eliminar os fatores que tumultuam o comportamento dos dados e, em consequência, elevar a representatividade ao ponto ideal.

Como bem afirma W. Jockran, "qualquer amostragem já realizada é, potencialmente, uma orientação para futuras amostragens melhoradas, pelos dados que fornece a respeito dos valores médios, dos desvios-padrões, da natureza da variabilidade das principais medidas e do custo de obtenção dos dados".

Exposta a metodologia, importa esclarecer o mecanismo de coleta e seleção das notas fiscais. O Decreto 62.158, de 19 de janeiro do corrente ano, referindo-se à segunda via da nota, dispõe que: art. 1º, II - a) "no caso de remessa por vias internas, será entregue diretamente pelo emitente à Agência Municipal de Estatística da Fundação Instituto Brasileiro de Geografia e Estatística, da jurisdição do seu domicílio, até o dia 10 de cada mês subsequente ao da emissão".

As Agências referidas como arrecadadoras da nota são administradas pela Inspeção Regional de Estatística, com a qual foi estabelecido o critério para o encaminhamento das notas ao Departamento de Estatística.

Tendo em vista a existência de duas subamostras, A e B, e considerando ainda o interêsse de aproveitar a procedência das notas, a fim de se estudar o fluxo, estabeleceu-se que a Inspeção Regional deveria enviar ao Departamento de Estatística o material já selecionado nos dois grandes grupos, isto é, um com valô-

res até NCr\$4 000 e outro com valores superiores.

A seleção das notas segundo as duas subamostras, feita no órgão coletor, foi uma boa solução para reduzir o trabalho no órgão apurador, que ficou, assim, liberada da penosa tarefa de seleção de grande volume de notas. Além disso, a cooperação da Inspeção Regional permitiu ainda a regionalização, pois a seleção se faz em função de cada Agência de Itinerância. Embora as áreas da A.I. tenham sido determinadas por critérios administrativos, de certa forma representam regiões mais ou menos homogêneas do ponto de vista geoeconômico. Dessa forma, o Departamento de Estatística recebe tôdas as notas com valores superiores a NCr\$4 000 e 10% das de valor inferior.

Tendo em vista o processamento mecânico dos dados, foi estabelecido um esquema de codificação das notas, visando apurar os seguintes dados: volume físico, valor, Unidade Federada de destino e área de procedência, segundo classes e subclasses de mercadorias, prevendo-se a possibilidade futura de oferecer resultados por mercadoria.

Durante a fase de codificação, está programada a crítica, indispensável em face da possibilidade de erro no preenchimento da nota. Aliás, experiência realizada no início do trabalho indica o acerto da medida saneadora.

Infelizmente, o pouco tempo de trabalho não nos permite oferecer resultados da experiência. Todavia, submetemos ao exame crítico da Conferência o nosso plano e a orientação metodológica, na certeza de merecer sugestões dos ilustres participantes do conclave.

O estudo que o Dr. Moysés Kessel apresentou constituirá, por certo, matéria do mais alto interêsse, pois seu autor aprecia o tema do ponto de vista de suas experiências e da metodologia.

Esta é a comunicação que o Departamento de Estatística de São Paulo tem a satisfação de oferecer à la. Conferência Nacional de Estatística.