

REVISTA BRASILEIRA DE ESTATÍSTICA

Órgão oficial do IBGE
e Sociedade Brasileira de Estatística

Redação:

Av. Franklin Roosevelt, 166 — ZC-39
Rio de Janeiro, RJ — Brasil — Tel: 252-3605

Diretor responsável:

Amaro da Costa Monteiro

Secretário:

Mário Fernandes Paulo

A Revista não se responsabiliza
pelos conceitos emitidos
em artigos assinados

Preço:

assinatura anual: Cr\$ 36,00
número avulso: Cr\$ 10,00

Vendas:

Av. Franklin Roosevelt, 146-A — loja B
Tel: 242-7142

SUMÁRIO

Artigos

- Considerações teóricas sobre o conceito de indicador social: uma proposta de trabalho
— Tereza Cristina Nascimento A. Costa 167
- Análise de correspondências: uma introdução
— Renato Galvão Flores Junior 177
- Tendência da mortalidade por causas no Brasil
— João Yunes, Jorge Somensi e Vera Shirley C. Ronchezel 257
- Aplicação de modelos multidimensionais à agricultura
— Leila Maia 281
- Estrutura do parque cafeeiro paranaense (safra 1971/72)
Orlando Sá Leite 299

Transcrições

- Organização e métodos do sistema dual de registros na Índia
— Joan Lingner e H. Bradley Wells 309

Noticiário

- 37º aniversário do IBGE 349
— Instituto Interamericano de Estatística 351

Bibliografia

- Pesquisa especial de bovinos 353
— Censo dos Serviços — Estado do Amazonas 354
— Censo Agropecuário — Paraná 355
— Boletim Informativo 356
— Publicações editadas pelos órgãos de Estatística do IBGE no trimestre abril-junho de 1975 356
— Censo dos Serviços: Rondônia-Roraima-Amapá, Maranhão, Ceará, Rio Grande do Norte, Alagoas, Sergipe, Espírito Santo, Santa Catarina e Distrito Federal 361

R. bras. Estat.	Rio de Janeiro	v. 36	n.º 142	p. 165 a 364	abr /jun. 1975
-----------------	----------------	-------	---------	--------------	----------------

Revista Brasileira de Estatística. v. 1- n. 1- jan /mar.
1940- Rio de Janeiro, Fundação Instituto Brasileiro de Geografia
e Estatística. Diretoria Técnica. Departamento de Divulgação Estatística,
1940-

v. 27cm Trimestral

Antiga estrutura deste órgão: Brasil. Conselho Nacional de Estatística Diretoria de Documentação e Divulgação e Instituto Brasileiro de Estatística. Departamento de Divulgação Estatística.

1. Estatística — Periódicos I. Fundação Instituto Brasileiro de Geografia e Estatística. Diretoria Técnica. Departamento de Divulgação Estatística, ed.

CDD 3105



CDU 31:05(81)

CONSIDERAÇÕES TEÓRICAS SOBRE O CONCEITO DE INDICADOR SOCIAL: UMA PROPOSTA DE TRABALHO

Tereza Cristina Nascimento A. Costa

Com este trabalho pretendemos levantar algumas questões relevantes quanto ao conceito de Indicadores Sociais que permitam chegar a um conceito de trabalho que possa orientar o Projeto Indicadores Sociais no âmbito do IBGE.

Consideraremos as seguintes questões:

1. Qual a especificidade do chamado movimento de Indicadores Sociais?
2. Os conceitos de Indicador e de Social
3. Definições de Indicadores Sociais
4. Uma opção de trabalho

A ESPECIFICIDADE DOS INDICADORES SOCIAIS

A análise histórica do desenvolvimento das sociedades mostra ter sido sempre o conhecimento dos especialistas utilizado pelo poder público. Igualmente uma história das ciências mostrará que há épocas e sistemas políticos que usam esta ou aquela ciência como instrumento de controle social. Isto tem se verificado com o Direito, a Biologia, a Economia e outras ciências.

Esta relação tem sido bastante intensa principalmente com as ciências sociais.

O que seria novo ou específico é a crescente presença da operacionalização que caracteriza as ciências sociais no momento em que o controle social tende a se expressar por meio de planejamento governamental.

Parece-nos também que há uma crescente consciência de que a abordagem predominantemente econômica do planejamento não daria conta por si só do real a ser controlado e planejado, verificando-se portanto uma tendência à incorporação de outras abordagens numa postura multidisciplinar.

A utilização de indicadores sociais parece-nos que deve ser pensada no contexto acima mencionado de uma relação entre conhecimento e controle social¹. Esta relação assume diversas naturezas conforme o tipo de conhecimento utilizado e também conforme as posições teóricas diversas passíveis de coexistirem em determinada ciência. Os vários conceitos de social que discutimos adiante, neste trabalho, ilustram o que acabamos de afirmar.

O SOCIAL

Qual seria o sentido do termo social no conceito de Indicadores Sociais?

Encontramos várias concepções de social tanto na literatura sobre Indicadores Sociais quanto em documentos de planejamento governamental. O termo social aparece definido de maneira residual ou mesmo negativamente, isto é, social é o fenômeno não econômico. Mais precisamente o social é pensado como fornecendo variáveis intervenientes entre as variáveis econômicas e o resultado que se pretende atingir. É desta maneira que o social vem sendo crescentemente incorporado ao planejamento governamental já que “a experiência logo evidenciou que, para alcançar o objetivo econômico de crescimento auto-sustentado da produção por habitante era necessário levar em consideração fatores não econômicos... A necessidade de incorporar o elemento “social” na planificação global chegou a se tornar um lugar comum nas resoluções dos organismos intergovernamentais”². Encontramos por exemplo esta consideração no I Plano Nacional de Desenvolvimento onde se afirma que “para conferir auto-sustentação e caráter integrado ao processo, o desenvolvimento pressupõe: transformação social para modernizar as instituições, acelerar o crescimento, distribuir melhor a renda e manter uma sociedade aberta”.

Encontramos também o social definido como sinônimo de bem-estar (welfare). Há toda uma série de trabalhos nesta linha como o de Allard, porém a definição de indicador social que torna meridiana

1 Ver a este respeito: DION. *Sociologie et ideologie dominante dans l'oeuvre de Le Play et Durkheim* La Pensée (158) adut 1971

2 BOLETIM ECONOMICO DE AMERICA LATINA. Santiago de Chile, CEPAL, v. 11, n. 1, 1966, p. 46

esta conotação de social é “um indicador *social*, tal como o termo é definido aqui, pode ser definido com uma estatística de interesse normativo direto,... É uma medida direta de bem-estar...” (Toward a social report) — US 1969:97³.

Esta abordagem do social denota uma perspectiva consensual da sociedade, sendo possivelmente vista como um somatório de indivíduos cujo bem-estar é somado para fornecer o bem-estar social. Quando tomamos social por bem-estar devemos explicitar a referência ao sistema de valores que orienta esta postura. Parece-nos que esta supõe uma sociedade sem conflito, onde há consenso sobre objetivos e valores e suas implicações. A sociedade é pensada como um todo homogêneo e não como um campo hierarquizado onde há valores em conflito. A definição de bem-estar é cultural, histórica e específica de determinada sociedade.

O CONCEITO DE INDICADOR

O termo indicador é sempre utilizado pela bibliografia sobre Indicadores Sociais justaposto ao termo social sem que mereça reflexão específica. Isto faz com que não se vincule os trabalhos de Indicadores Sociais à tradição de pesquisa sociológica.

Quando discutimos o conceito de indicador estamos reivindicando uma atitude que não reduza o trabalho de pesquisa a enumeração de definições operacionais sem que haja um trabalho de elaboração teórica que defina a fecundidade dos conceitos.

Nesta linha diríamos que o trabalho científico consiste na formulação de sistemas conceituados dotados de pertinência teórica.

Quando trabalhamos com indicadores temos que enfrentar além do problema fundamental mencionado acima e que diz respeito a elaboração teórica os seguintes:

- 1) Qual a relação entre o indicador e o conceito?
— Mensuração fracionada
- 2) Retradução de dados
Uso de unidades “formais” x “reais”
Medidas agregadas x globais

RELAÇÃO ENTRE O INDICADOR E O CONCEITO

O trabalho de construção de indicadores supõe de certa maneira, que o indicador esteja de um modo ótimo referido ao conceito. A for-

³ Sobre as definições correntes de Indicadores Sociais ver o trabalho de L. Parga Nina — Fundação IBGE, nov. 1972. Mimeografado
O trabalho de Allard está baseado no levantamento de necessidades que classifica em “having”, “loving” e “being” ou seja, necessidades materiais, integração social e auto-realização. Para uma crítica ao conceito de necessidades e de seus postulados assim como sua relação com consumo — ver BAUDRILLARD, Jean La Genèse Ideologique des Besoins *Lahiers Internationaux de Sociologie*, Paris 47, 1969

mulação clássica que ilustra este ponto de vista é a de Durkheim: “A solidariedade social é um fenômeno moral que, por ela mesma, não se presta à observação exata e sobretudo à medida. Para proceder tanto à classificação quanto à comparação é preciso portanto, substituir ao fato interno que nos escapa um fato exterior que o simboliza e estudar o primeiro através do segundo. Com efeito, onde a solidariedade social existe, apesar do seu caráter imaterial, ela não fica num estado de pura potência porém manifesta sua presença por efeitos sensíveis”⁴.

O texto de Durkheim nos indica alguns elementos presentes em reflexões teóricas sobre Indicadores Sociais:

→ o indicador permite o estudo do conceito

→ o indicador é mensurável

Todo problema consiste em determinar um indicador que efetivamente “simbolize” o conceito. É difícil que para tanto possamos utilizar apenas um único indicador, uma alternativa seria utilizar vários indicadores porém, isto é, eficiente só na medida em que as várias dimensões do conceito são cobertas.

Uma das tendências nos projetos de Indicadores Sociais tem sido a de combinar os vários indicadores construídos em um só índice, arriscando com isto perder informações. O estudo de nível de vida da UNRISD leva este problema em conta afirmando seus técnicos que: “o índice global deve permitir entretanto, estudar a relação estabelecida entre os vários indicadores pois esta é fundamental. No nosso caso nos parece que o importante não é o nível atingido em cada indicador dos sete componentes do nível de vida, mas a forma do espectro que eles delinham”⁵.

O texto indica igualmente uma preocupação com a mensuração do conceito. Acreditamos ser de duas ordens a ênfase quantitativa nos projetos Indicadores Sociais. A primeira diria respeito a utilização no planejamento governamental dos Indicadores Sociais. O planejamento enfatizaria o caráter operacional dos Indicadores Sociais privilegiando estatísticas sociais com vistas a implementação de programas. A segunda estaria referida a relação entre mensuração e construção de teoria. Esta seria, segundo Coleman, extremamente fecunda visto que o “poder de uma teoria de fornecer deduções precisas e numerosas depende de sua habilidade em efetuar transformações sobre os dados. Se estes dados estão em forma de números e mantêm suas propriedades de números após as transformações então as poderosas transformações da álgebra, cálculo, álgebra matricial podem ser empregadas”⁶. (Coleman 1964:56)

4 DURKHEIM *De la division du Travail Social* 1967, p 286

5 VERGÉS, Pierre. Une interpretation structurale des composants du niveau de vie. *Developpement et Civilisations*, Paris (36): 42 dec 1968

6 COLEMAN, James C. *Introduction to Mathematical Sociology* Free Press, 1964. p. 56

Queremos entretanto enfatizar que nos parece que o conceito de Indicadores Sociais não deve estar vinculado a uma perspectiva quantitativa que limite a pesquisa sociológica apenas ao que é mensurável, mas ao que é relevante. Cremos que muitas vezes teremos que optar entre uma medida sofisticada de um conceito menos fecundo e uma medida bruta de um conceito relevante. Há ainda muito que se fazer no sentido de tornar o que é relevante mensurável.

Outros problemas que devem ser ao menos mencionados são: a retradução que se faz necessária operar nos dados que não foram levantados para o objeto de estudo em questão; um exemplo são os trabalhos baseados em registro administrativo sobre educação, casamentos, conflitos de trabalho, suicídios e outros em que conceitos são elaborados a partir dos objetivos da instituição competente e não das necessidades específicas do pesquisador. A respeito desta problemática diz Bordieu "é suficiente ter tentado uma vez submeter à análise secundária um material recolhido em função de uma outra problemática, por mais neutra que seja em aparência, para saber que os dados, os mais teóricos não poderiam responder completa e adequadamente a questões para as quais e pelas quais não foram construídos. Não se trata de contestar por princípio a validade de utilização de um material secundário, mas de explicitar as condições epistemológicas deste trabalho de retradução que se refere sempre a fatos construídos (bem ou mal) e a dados. Tal trabalho de retradução, o qual Durkheim exemplifica com sua obra "O Suicídio", poderia mesmo ser o melhor incentivo à vigilância epistemológica na medida em que este exige uma explicitação metódica da problemática e dos princípios de construção do objeto que estão presentes tanto no material quanto no novo tratamento a que este é submetido"⁷.

Esta retradução nos fornece exatamente a indicação precisa de que não há sentido nos fatos sem teoria.

Outro problema é a utilização de dados levantados para determinadas unidades, por exemplo administrativas e/ou geográficas e que são tomadas por unidades sociologicamente relevantes. É usual também a agregação de dados de determinadas unidades gerando, por esta operação, grupos e instituições em vez de obter-se medidas globais próprias do grupo ou da coletividade⁸.

DEFINIÇÕES CORRENTES DE INDICADORES SOCIAIS

Podemos relacionar as definições correntes de Indicadores Sociais com as concepções do social que mencionamos anteriormente.

7 BOURDIEU, P.; CHAMBOREDON, J. C. & PASSERON, J. C. *Le Métier de Sociologue*. Paris, Mouton Bordas, 1968. p. 63

8 Um bom levantamento das posições assumidas quanto a validade do uso de estatísticas oficiais em Sociologia é de HINDESS, Barry. *The use of Official Statistics in Sociology* — A Critique of Positivism and Ethno Methodology. New York, Macmillan, 1973

Não iremos enumerar as várias definições possíveis de Indicadores Sociais mas classificá-las segundo um critério: o de conceituação do social tal como abordado anteriormente. Duas conceituações de Indicadores Sociais podem ser mencionadas e discutidas e que explicitam noções diferentes do social.

“Um indicador social é — uma medida de fenômenos sociais que são transeconômicos. O Indicador Social é normativo e integrado num sistema de informação consistente”⁹.

“Um indicador social, pode ser entendido como uma definição operacional ou parte de uma definição operacional de qualquer um dos conceitos centrais à geração de um sistema de informação descritivo de um sistema social”¹⁰.

Entre as definições possíveis de Indicadores Sociais já mencionamos a do “Toward a Social Report” que considera indicador social uma medida direta de bem-estar. A definição proposta por Cazes adota esta perspectiva enfatizando, entretanto, que é essencial a referência ou melhor a explicitação do sistema de valores que definiria o bem-estar.

Parece-nos que os vários estudos de qualidade de vida e de nível de vida se enquadram nesta abordagem normativa¹¹.

Esta definição relaciona Indicadores Sociais a fenômenos sociais transeconômicos, assumindo assim que a abordagem econômica não seria suficiente para a análise do social. É importante lembrar que muitas vezes o apelo a uma abordagem sociológica tem sido apenas ritual uma vez que se procura levantar Indicadores Sociais a imagem dos Indicadores Econômicos.

A segunda definição identifica social a sistema social, assume também que o indicador social é uma medida já que o identifica a uma definição operacional.

Parece-nos que as definições tem implicitamente dois tipos de orientação respectivamente.

A definição normativa indicaria uma preocupação maior em iluminar a escolha do planejador, em fixar e avaliar quantitativamente as finalidades e prioridades e carências da sociedade face a determinados objetivos. A outra definição tipo sistema social está centrada em apreender e explicar o sistema social em suas estruturas e princípios organizatórios.

9 CAZES, Bernard The development of social indicators: a survey In: SOCIAL Indicators and Social Policy, Ed Shonfield and Shaw, SSRC, 1972, p. 14

10 CARSLILE, Elaine The conceptual structure of social indicators In: SOCIAL Indicators and Social Policy, Ed Shonfield and Shaw, SSRC, 1972 p 25

11 Alguns estudos de Nível de Vida explicitam o sistema de valores que está orientando o trabalho tornando-o assim mais “científico”. Um exemplo é o Level of Living Study do governo da Noruega no qual é enfatizado ter sido a iniolativa de fazer o estudo de autoridades políticas e que o fenômeno da desigualdade é visto como uma característica de sociedades ricas e da Noruega em particular. O estudo considera a desigualdade e a existência de grupos de baixa renda um problema. Ver RINGEN. *An Introduction to the Level of Living Study* 1973. Mimeografado

Acreditamos que o indicador social deve estar sempre referido a sistema social e que a segunda definição não é conflitante com a primeira mas vai além dela ¹².

UMA OPÇÃO DE TRABALHO

Diante das considerações feitas devemos assumir uma posição privilegiando como opção de trabalho determinada linha de pensamento sobre Indicadores Sociais.

Para fazer esta opção cremos que haveriam alguns pressupostos comuns a uma série de Projetos de Indicadores Sociais seja na área governamental, acadêmica ou da comunidade internacional que gostaríamos de expor e nos situarmos com relação a eles:

1. Indicadores são normativos
2. Indicadores permitirão o estabelecimento de objetivos e prioridades sociais
3. A falta de informação é que explicaria a inexistência de determinadas políticas, ou a irracionalidade e insucesso de outras

Quando anteriormente criticamos o caráter normativo de determinadas definições de Indicadores Sociais estávamos nos referindo a definições que identificam indicador social à medição de bem-estar social. Creio que devemos deixar claro que consideramos que a atividade do cientista social e do planejador está referida sempre a um sistema de valores, e isto se expressa por exemplo, na medida em que estes cientistas fazem determinadas perguntas e não outras gerando informação sobre certos sistemas e não sobre outros ¹³.

A perspectiva assumida por alguns teóricos de projetos de Indicadores Sociais é de que as informações geradas pelos Indicadores Sociais permitirão o estabelecimento de objetivos e prioridades sociais; achamos importante relativizar esta proposição afirmando que a definição destes objetivos e prioridades é realizada pelos diversos grupos sociais que podem utilizar o conhecimento dos cientistas sociais para o estabelecimento de políticas.

O pressuposto seguinte tem relação com o anterior na medida em que acreditamos não ser a falta de informação sobre determinados pro-

12 A nossa posição pode ser expressa da seguinte maneira: "Os critérios quantitativos para fixar fundos que se destinarão a programas sociais ou taxas de aumento nos componentes de nível de vida são menos importantes que a compreensão do funcionamento dos programas em uma determinada estrutura social e sua potencialidade para modificar esta estrutura de modo mais adequado ao desenvolvimento". BOLETIM ECONOMICO DE AMERICA LATINA Santiago de Chile, CEPAL, v. 11, n. 1, 1966, p. 50

13 Consultar a respeito WEBER *The Methodology of the Social Sciences* Free Press, 1969

blemas sociais a razão da sua persistência mas sim o conflito de ideologias¹⁴.

Discutidos estes pressupostos, assumidos ou não pelos cientistas sociais que trabalham na área de Indicadores Sociais, é possível adotar uma entre as várias definições de Indicadores Sociais.

No decorrer do trabalho desenvolvemos o quadro de referência que geraria uma definição adequada de Indicador Social. Deste quadro destacam-se alguns pontos:

- Social está referido a sistema social
- Há aspectos de funcionalidade e disfuncionalidade no levantamento de indicadores
- Não há sentido no indicador social sem construção de teoria
- A utilização de metodologias diversas, por exemplo, de caráter estatístico e etnográfico, é fundamental para a construção de um sistema de indicadores¹⁵.

Diríamos portanto que o conceito de *Indicador Social* se refere a dados, inclusive estatísticos, que estariam referidos a conceitos centrais à geração de um sistema de informação sobre sistemas sociais. Estes conceitos centrais são percebidos como componentes de um modelo de sistema social. Este modelo poderíamos definir como “concepções de processos sociais que podem ser formulados verbalmente, logicamente e matematicamente ou por simulação...¹⁶.

Tem-se afirmado que a teoria sociológica ainda não forneceu uma teoria geral aplicável a sistemas sociais. Ainda que se aceite esta posição, alguns modelos têm sido elaborados para subsistemas sociais para determinados princípios estruturais dos sistemas sociais¹⁷.

Acreditamos portanto, ser possível adotar esta definição para continuar os trabalhos do Grupo Projeto Indicadores Sociais.

O Grupo conforme a orientação acima tem desenvolvido trabalhos na área de educação, fecundidade, renda, mão-de-obra, estudos urban

14 Sobre as implicações políticas de um Projeto de Indicadores Sociais — Formulação e avaliação de questões, controle e uso da informação, impacto no sistema social do desenvolvimento, etc ver HENRIOT. *Political questions about Social Indicators* *West African Political Quarterly*, 23 (2) June 1970 Ver também SISTEMAS de informação e indicadores políticos de emprego Rio de Janeiro, IPEA/IBGE, ago 1974

15 Acreditamos ser correta a afirmação de Boudon de que “Não vemos como a sociologia poderia chegar a uma unificação de seus métodos sem renunciar a diversidade de objetos que ela se propõe atualmente” — BOUDON. *Métodos quantitativos em ciências sociais* Rio de Janeiro, Vozes, 1971 p 116

16 LAND, Kenneth C. On the definition of Social Indicators *American Sociologist*, 6 (1971), p 322-5

17 Ver STONE, Richard *Demographic Accounting and Model Building* Paris, OECD, 1967 Ver também BLAU, Peter & DUCAN, Otis D *The American Occupational Structure* New York, Wiley, 1967 Parece-me que as obras de Durkheim, Weber e Marx oferecem uma teoria geral sobre sistemas sociais Recentemente Parsons vem tentando elaborar tal teoria. Consultar a este respeito THEORIES of Society, Free Press, 1970

O Grupo tem adotado modelos de distribuição de atividades ou modelos de desigualdade social, que constituem uma área de pesquisa tradicional das ciências sociais. A pesquisa de mobilidade social em desenvolvimento no Grupo, acreditamos, trará novas contribuições para o estudo do desenvolvimento sócio-econômico da sociedade brasileira, permitindo o levantamento de indicadores de estratificação social, mobilidade ocupacional, etc.

Outra orientação de trabalho adotada consistiu na adoção das sugestões da ONU para a construção de Indicadores Sociais contidas no documento — “A System of Demographic and Social Statistics and its link with the System of National Economic Accounts”, 1972. Esta proposta da ONU parece-nos estar enquadrada, numa abordagem que considera ser a construção dos Indicadores Sociais fundamental para “o estudo das condições sociais e da mudança social em seus vários aspectos”; relaciona também indicador social à medição de bem-estar. O documento da ONU se propõe “a descrever métodos para o estabelecimento de modelos da progressão de seres humanos através do sistema social ou de alguma de suas partes” (parágrafo 7.1).

Sugere ainda a construção de dois tipos de indicadores sociais — simples e complexos. A diferença entre os dois tipos estaria no tratamento analítico mais ou menos sofisticado. Podemos dizer que os indicadores que sugere a ONU estão em sua maior parte aquém das possibilidades do sistema de informação do IBGE em determinados campos como População, População Economicamente Ativa e Educação — indicadores já construídos pelo GPIS a nível de Brasil para 1970.

ANÁLISE DE CORRESPONDÊNCIAS: UMA INTRODUÇÃO

Renato Galvão Flores Jr.

Estadístico

SUMÁRIO

NOTA PRELIMINAR

PARTE I

Introdução

Cap 1 — Noções sobre tabelas de contingência

- 1.1 — Principais definições*
- 1.2 — Tabelas de descrição lógica*
- 1.3 — Outros tipos de tabelas de contingência*
- 1.4 — Transformações em tabelas de contingências*
- 1.5 — A análise estatística clássica das tabelas de contingências*

Cap 2 — O Método de análise de correspondência

- 2.1 — Nuvens de pontos e eixos fatoriais*
- 2.2 — Fatores e fórmulas*
- 2.3 — A métrica do Qui-quadrado*
- 2.4 — Considerações finais*

Cap 3 — Alguns exemplos ilustrativos

Cap. 4 — Algumas extensões e resultados particulares

- 4.1 — Efeito Guttman*
- 4.2 — Testes sobre os resultados em correspondências*

Cap 5 — Uma visão das aplicações da análise de correspondência

5.1 — Correspondências e teoria da informação

5.2 — Correspondências e classificação hierárquica

5.3 — Correspondências e consistência estatística de dados

5.4 — "Scaling", "parametric representation" e correspondências

5.5 — Grafos e correspondências

5.6 — A avaliação estatística de sistemas de computação e a análise de correspondências

PARTE II

Introdução

Cap 1 — A posição social das ocupações

Cap 2 — Um estudo por correspondências das Microrregiões do Estado do Espírito Santo

BIBLIOGRAFIA

NOTA PRELIMINAR

Nestas páginas discute-se o método de análise de correspondências. Trata-se da técnica estatística, voltada para o tratamento de grande massa de dados via computador, que vem sendo utilizada no Centro de Informática do IBGE. As rotinas para perfazê-la encontram-se integradas no sistema de análise de dados aqui desenvolvidos — o SADE.

A estrutura do trabalho é simples e a linguagem é bastante informal na maioria das vezes. Não há pretensão de originalidade ou refinamento matemático e sim de disseminar ao maior número possível de pesquisadores e interessados as idéias mínimas para compreender e saber utilizar o método em questão. De início encontram-se os conceitos, fórmulas e implicações, ou seja, a teoria; tendo-se procurado não obscurecer os fatos com manipulações formais minuciosas. Poucos são os resultados demonstrados rigorosamente, embora sempre se faça menção à fonte original. Numa segunda parte apresenta-se dois exemplos concretos obtidos dos arquivos de dados do Centro de Informática e executados pela nossa implementação. Deseja-se que constituam uma ilustração interessante e didática (não houve pretensão de pesquisa) do anteriormente exposto. No início de cada parte existe um capítulo introdutório visando dar uma noção geral do que vai se seguir.

Para o autor isto é um dos resultados de um contato de dez meses com o método. Muita coisa foi antes exposta e debatida em um seminário desenvolvido em meados de 1974. Na realidade aqui estão também representados o auxílio e cooperação que recebi dos colegas, desde que iniciou a leitura dos alentados tomos de "J-P Benzécri et Collaborateurs". Gostaria de agradecer a Sergio Ivan Roschke que for-

neceu muitos “outputs” de correspondências obtidos via um programa particular, e que foram exemplos muito úteis no início do estudo. Além dele, a Milton Corrêa Filho que como responsável pela confecção da rotina muito contribuiu na ilustração de detalhes de cunho informático. Os programadores Edson Buriche Coutinho, Oscar Luís de Farias e Pedro Vieira do Nascimento foram sempre companheiros agradáveis e competentes. Porém, de maneira muito especial, o autor agradece a Luiz Carlos Gomes, cuja personalidade inspiradora e cativante fá-lo genuíno mentor do trabalho realizado e dos frutos que porventura produza.

PARTE I

INTRODUÇÃO

Basta um simples folhear de um dos volumes dos congressos internacionais de análise multivariada, como os editados por Krishnaiah [13], [14], ou um rápido percorrer do índice de matérias dos últimos números do JASA ou qualquer outra publicação estatística atual, para se notar que os métodos da análise multivariada encontram-se em franca e, de certo modo, desordenada expansão. Origina-se isso talvez, do irrestrito uso que se faz de suas idéias básicas no vastíssimo campo das ciências sociais e humanas, bem como no das ciências físicas, gerando assim um número imenso de técnicas afins a menos de pequenas variantes — só em psicometria, por exemplo, existem perto de cinquenta variações de análise fatorial — e de particulares pontos de vista e justificativas, apropriadas à especial disciplina do pesquisador. Ao mesmo tempo, e por isso mesmo, os resultados teóricos dentro do espírito da estatística matemática, que se reflete numa obra clássica como a de T. W. Anderson [1], ou nos artigos sobre o assunto publicados no “Annals of Mathematical Statistics”, são impotentes no sentido de unificar e justificar, assentando sobre bases formais precisas os diversos resultados obtidos empiricamente. Não fora porque pela sua natureza os teoremas da estatística matemática necessitam, na maioria das vezes, assumir excessivo número de hipóteses *a priori* sobre os vetores/observação, hipóteses essas violadas em muitos casos práticos.

Dentro pois desse atual estado da arte em análise multivariada, é que deve ser encarada a análise de correspondências, parte de um trabalho de pesquisa desenvolvido pelo Laboratório de Estatística da Universidade de Paris VI, chefiado pelo professor Jean-Paul Benzécri. Como suas irmãs, a análise de correspondências nasceu e é apreendida muito mais pelo “feeling” desenvolvido através de um contacto

íntimo com a massa de dados do que com a busca por resultados gerais e perfeitos matematicamente. Prescinde ela de suposições *a priori* sobre a natureza da distribuição das observações e possui todo o espírito do que J. W. Tukey chama de uma técnica exploratória ([22], [23]), ou seja: tratamentos iniciais a que se deve submeter os dados a fim de se tentar extrair estruturas ou relações inerentes ao próprio conjunto, independentes de qualquer asserção preliminar que se tenha feito sobre eles. Isso é exatamente o que J. P. Benzécri, juntamente com vários estatísticos de renome da atualidade, chama de *análise de dados*.

Otvei a análise de correspondências apreciável sucesso nas pesquisas em que foi utilizada, como está fartamente ilustrado em [2], e, cremos, grande parte desse bom resultado é devido ao fato de que ela unifica de forma muito feliz os diversos métodos da análise fatorial e suas derivações, como a análise de aglomerados, podendo substituí-los *com eficácia* na maioria das vezes. Além do mais, por possuir, como se verá, uma estrutura mais livre e desembaraçada de hipóteses a sua aplicação se estende a casos onde a análise fatorial, ou algo afim, *seria impraticável*. Desse modo o seu espectro de aplicações é consideravelmente largo com relação aos outros métodos multivariados.

Nessa primeira parte, procura-se expor no Capítulo 1 os conceitos fundamentais relativos aos cruzamentos ou tabelas de contingência, uma vez que são elas o ponto de partida do método. No Capítulo 2 explica-se o próprio método procurando não carregar exageradamente na parte matemática. Apenas conhecimentos básicos de álgebra linear e estatística são invocados, a fim de dar um mínimo de suporte teórico à discussão. Após isso, dedica-se um capítulo inteiro à discussão de exemplos que ilustrem a teoria. O Capítulo 4 representa uma digressão mais avançada e pode ser omitido pelo leitor mais apressado ou com menos preocupações formais. Finalmente, o Capítulo 5 é uma pequena visão geral de onde se tem aplicado a análise de correspondências, comentando-se com um pouco mais de detalhe duas ou três aplicações específicas.

Capítulo 1

NOÇÕES SOBRE TABELAS DE CONTINGÊNCIA

Uma vez que a análise de correspondências repousa fortemente sobre os conceitos relativos a tabelas cruzadas, dedicaremos esse capítulo a um estudo básico de tais conceitos e de algumas idéias que nos serão importantes.

1.1 Principais definições

1.1.1 A tabela de contingência nada mais é do que o cruzamento de várias categorias, definindo assim um conjunto de tipos ou celas, onde se conta o número de indivíduos em cada cela. De uma maneira mais formal, uma tabela de contingência a duas dimensões a $|I|$ linhas e $|J|$ colunas é simplesmente o resultado da amostragem de um vetor aleatório bidimensional (X_1, X_2) . Tal vetor deve ser discreto (=categórico) ou ter sido previamente discretizado (=categorizado) em um número finito de ocorrências possíveis através uma classificação. Ao conjunto de valores de X_1 , convencionou-se chamar conjunto de linhas e o denotaremos sempre por I , quanto ao de X_2 , será o conjunto J de colunas. Assim a tabela a duas dimensões é uma matriz $|I| \times |J|$ onde em cada cela $(i, j) \in I \times J$ coloca-se o número de vezes que o vetor aleatório (X_1, X_2) caiu nessa classificação. Costuma-se representar a entrada da tabela correspondente à cela (i, j) pelo símbolo n_{ij}

Exemplos:

- i) Consideremos uma pesquisa onde se seleciona uma amostra de pessoas de vários municípios do Brasil e se levanta as respectivas ocupações. Colocando-se o conjunto dos municípios nas linhas e o das ocupações em colunas, cada entrada (i, j) da tabela nos fornece o número de habitantes do particular município i que tem a ocupação j .
- ii) Deseja-se medir a “reação” de um grupo social a “ações criminosas”, para isso, entre outras coisas, entrega-se a uma amostra de pessoas desse grupo uma lista de tais delitos e se solicita que a pessoa dê uma nota de 1 a 10 proporcional à gravidade que ela atribui ao delito. Uma tabela de contingência forma-se naturalmente ao se cruzar os delitos com notas, colocando-se em cada interseção (i, j) o número de indivíduos que deu a nota j ao delito i .
- iii) Submete-se um grupo de pessoas a um questionário formado de 6 perguntas, das quais quatro possuem como respostas sim ou não, e as outras duas são perguntas que enquadram o entrevistado em uma dentre várias classes. Colocando-se nas linhas o grupo de indivíduos e nas colunas as diversas respostas possíveis, a saber: sim e não para as quatro perguntas, e as várias classes dadas pelas outras duas; teremos uma tabela de contingência onde na cela (i, j) opõe-se o número 1 ou 0, conforme a resposta do i -ésimo indivíduo tenha selecionado ou não a categoria j .
- iv) Voltando ao exemplo *ii* suponha-se que se cruze agora os indivíduos da amostra contra os delitos, representando em cada cela a nota que o indivíduo correspondente deu ao delito respectivo.
- v) Em estudos econômicos é comum cruzar-se as indústrias (I) de uma certa região com a lista de insumos (J) usados na região, colocando-se em cada cela (i, j) o valor, em cruzeiros, do insumo j que é consumido pela indústria i .

Conforme se pode notar, os três últimos exemplos são um pouco mais delicados quando se tenta enquadrá-los na definição dada no início. Na realidade, muitas vezes a matriz de cruzamentos apresenta valores não inteiros, ou que possuam uma interpretação artificial como “número de indivíduos em uma dada categoria”. Como, sob o ponto de vista de correspondências, isso não constitui entrave algum, passaremos a permitir tal flexibilidade para os nossos cruzamentos, sem se alongar sobre esse detalhe.

1.1.2 Embora se tenha ampliado o conceito de tabela de contingência, não exigindo mais que as entradas sejam resultados de contagem, não se deve entretanto, devido à natureza de método de correspondências, se apresentar um cruzamento onde as celas não sejam dimensionalmente homogêneas. É por causa disso que em *iv* se equalizou a contribuição dos insumos em cada estabelecimento através o seu valor em cruzeiros. O Exemplo 2 da Parte II fornece uma boa ilustração de um cruzamento onde as entradas são tornadas adimensionais devido a provirem de medidas em unidades diversas. Tal fato portanto, a que deve obedecer toda tabela que for ser submetida a uma correspondência, denomina-se usualmente *Princípio da homogeneidade dimensional*.

Outra condição a que uma tabela de contingência deve satisfazer é a de, teoricamente, esgotar o universo de ocorrências possíveis do vetor (X_1, X_2) . Resulta essa exigência em algo extremamente delicado e discutível, quando de estudos do tipo descrito em *ii*. Nesse caso por exemplo, assume-se de certa forma que o conjunto de delitos submetido às pessoas da amostra esgota o “universo de ações criminosas” para aquele grupo. Com isso em mente, deve-se ter extremo cuidado ao se interpretar o significado preciso dos dois universos representados pelas classificações da tabela em suas duas dimensões, a saber, as linhas e colunas; a má compreensão de qualquer um deles pode acarretar erros na interpretação dos resultados da análise. A esta segunda condição denomina-se comumente *Princípio da Exaustividade*.

1.2 Tabelas de descrição lógica

Diz-se que uma tabela de contingência a linhas I e colunas J é uma tabela de descrição lógica se e só se:

$\forall (i, j) \in I \times J, n_{ij} = 0$ ou 1 ; os valores 0 e 1 devendo ser interpretados como valores booleanos. Assim se $n_{ij} = 0$ (resp. 1) diz-se que o indivíduo i não possui a propriedade j (resp. i possui a propriedade j).

Um caso particular importante das tabelas de descrição lógica é o das que podem ser colocadas sob forma disjuntiva completa. Chama-se a isto quando o conjunto J (das propriedades) pode ser particionado em subconjuntos disjuntos J_K , de tal maneira que cada indivíduo i

possua uma e somente uma propriedade em cada J_K . Mais formalmente, se e só se:

$$\exists \{J_K\} K = 1, \dots, m \text{ tal que } J = \sum_{K=1}^M J_K$$

e

$$\forall i \in I, \forall J_K, \exists j \in J_K: n_{ij} = 1 \text{ e } j' \in J_K, j' \neq j \implies n_{ij'} = 0$$

O exemplo *iii* do parágrafo anterior constitui-se numa tabela disjuntiva completa. Na realidade, sempre que cada pergunta K , de um dado questionário, seja representada pelo seu conjunto de respostas possíveis J_K e que se se cruze o conjunto de indivíduos I com a união dos J_K da maneira descrita em *iii* teremos uma tabela de contingência sob forma disjuntiva completa. Para outro exemplo recomendamos o Exemplo 1 da Parte II.

1.3 Outros tipos de tabelas de contingência

Conforme o particular desejo do pesquisador, casos especiais de tabelas cruzadas foram e são desenvolvidos. No nosso estudo ficamos com os acima descritos, acrescentando apenas que quando o conjunto das linhas é idêntico ao das colunas costuma-se dizer que a tabela é simétrica. Resta entretanto fazer uma observação sobre as tabelas a mais de duas dimensões, ou múltiplas: nada mais são, formalmente, do que funções numéricas que partem de um produto cartesiano a mais de dois fatores. Isto é, a tipologia da população passa a ser gerada pelo cruzamento de 3 ou mais ocorrências.

Exemplo: Estudando-se a evolução da configuração das ocupações J em um conjunto I de municípios ao longo de um certo número de instante de tempo (T), em cada *data* tira-se uma amostra de pessoas, de tamanho constante, e se cruza em $I \times J$ como habitualmente. O cruzamento final $I \times J \times T$ nos dá então exatamente a evolução do perfil da distribuição das ocupações pelos municípios.

1.4 Transformações em tabelas de contingências

Muitas vezes uma tabela de contingência como levantada na pesquisa necessita ser alterada conforme o tipo de estudo que se pretenda fazer com seus conjuntos marginais. Um mesmo cruzamento pode assim dar origem a vários outros, que submetidos à análise de correspondências permitirão diferentes saídas/interpretações. De outra feita tabelas de contingência aparentemente distintas, podem através determinadas transformações, mostrar-se equivalentes segundo o resultado de uma correspondência. Mostramos a seguir algumas das transformações mais usuais.

1.4.1 Desdobramento

Muitas vezes em uma tabela como a do exemplo *iv* de 1.1.1, em que as entradas são inteiros que variam em um intervalo limitado, é comum desdobrá-la da seguinte forma:

Sejam I e J as margens da tabela (pensar, como no exemplo citado, em I sendo um conjunto de indivíduos e J de delitos) e suponha-se que a nota máxima que um indivíduo possa atribuir a um delito seja M . Formemos uma nova tabela fazendo com que cada coluna da primeira dê origem a duas outras: uma igual a ela própria e outra em que cada entrada seja igual a M menos o valor anterior. Isso significa associar a cada indivíduo/delito duas notas: uma de gravidade, que é a nota dada pelo indivíduo e que estava na cela correspondente da tabela inicial, e outra de tolerância, igual à maior gravidade possível (M) menos a nota atribuída, e que preenche a nova cela criada.

Mais formalmente chamado de J^* o novo conjunto de colunas, J^* é a união disjunta de J , o conjunto anterior com J' , o conjunto das colunas/delitos com as notas de tolerância. Uma nova entrada qualquer n_{ij^*} será igual a n_{ij} caso $j^* \in J$, e será igual a $M - n_{ij}$ caso $j^* \in J'$ e j seja a "coluna de gravidade" associada ao mesmo delito.

Notemos que uma tabela de descrição lógica em forma disjuntiva completa, i e j , uma tabela provinda de um questionário, pode ser encarada como uma tabela desdobrada caso todas as perguntas do questionário só admitam duas respostas. Ela será assim o desdobramento de uma tabela com a metade de suas colunas, obtida associando-se a cada pergunta uma única resposta. Nesse caso, obviamente, $M=1$

1.4.2 Transformação em tabelas simétricas

Em muitos estudos, como por exemplo os de classificação, necessita-se suprimir uma das margens da tabela $I \times J$ transformando-a em outro cruzamento, v. g., $J \times J$.

Exemplo: Suponhamos que usando-se a tabela levantada em *v* de 1.1.1 se deseje obter uma classificação dos insumos conforme a contribuição às indústrias refletida pela tabela. Os algoritmos de classificação usando correspondências exigem a entrada de uma matriz simétrica definida sobre o conjunto a se classificar.

A transformação geralmente usada é a que resulta em associar a cada par de insumos (j' , j'') o somatório das contribuições de ambos a cada indústria examinada. Tal contribuição é geralmente definida

como $\frac{n_{ij'} \times n_{ij''}}{\sum_{j \in J} n_{ij}}$ para cada indústria i ; sendo então a entrada $s_{j'j''}$

da nova matriz o valor:

$$s_{j'j''} = \sum_{i \in I} \frac{n_{ij'} \times n_{ij''}}{\sum_{j \in J} n_{ij}}$$

Em muitos outros casos, não só de origem econômica, a forma de transformação acima descrita é utilizada.

As tabelas $I \times J$ em forma disjuntiva completa são também muitas vezes simetrizadas mediante a transformação:

$$t_{jj'} = \# \{i \in I \mid n_{ij} = n_{ij'} = 1\}$$

Exemplo: Explorando o exemplo do questionário, $t_{jj'}$ representará o número de indivíduos que pertencem simultaneamente às classes j e j' (ou responderam afirmativamente a j e j'). Obviamente se j e j' estão numa mesma pergunta, $t_{jj'}$ será zero pois a tabela é disjuntiva completa. (Em geral, $j, j' \in J_K \implies t_{jj'} = 0$).

1.4.3 Transformações de tabelas múltiplas

As tabelas múltiplas são muitas vezes bidimensionadas agregando-se adequadamente os fatores do produto cartesiano.

Exemplo: No caso citado em 1.3 poderíamos passar a analisar a nova tabela $L \times J$, onde $L = I \times T$, que nos dá para cada ocupação j a sua evolução ao longo do tempo no espaço dos municípios.

Existe no entanto uma transformação que representa uma tabela múltipla $I_1 \times I_2 \times \dots \times I_m$ como uma tabela disjuntiva completa a duas dimensões.

Baseia-se ela em uma analogia com a simetrização das tabelas disjuntivas completas:

Repare-se que o produto cartesiano $P = \prod_{K=1}^m J_K$ dos conjuntos/perguntas da tabela disjuntiva é simplesmente o conjunto de todas as séries de respostas possíveis que se pode obter com o dito questionário, pois cada $p \in P$ é uma m -pla de respostas/columnas j_k associadas às perguntas J_k , $1 \leq K \leq m$. Usando-se a tabela original pode-se extrair uma distribuição de probabilidade em P , que chamaremos f_P , dividindo-se para cada série de respostas o número de indivíduos que a escolheram pelo número total de indivíduos entrevistados.

Simetrizando-se a tabela pela maneira descrita em 1.4.2, i.e., a cada par de respostas associa-se o número de pessoas que o escolheu, podemos relacionar a distribuição f_P com as entradas $t_{jj'}$ da nova tabela. Basta notar que a probabilidade do conjunto das séries de respostas possíveis em que figuram j e j' é simplesmente o somatório dos valores de f_P em cada ponto desse conjunto, ou seja, é o quociente do número de indivíduos cuja série de respostas se enquadra nesse conjunto pelo número total de indivíduos, logo, multiplicando-se esse valor pelo cardinal de I , teremos o número de indivíduos que escolheram as respostas j e j' isto é o valor de $t_{jj'}$. Portanto demonstramos a igualdade:

$$t_{jj'} = |I| \times \sum_{p \in \{p \in P \mid \text{em } p \text{ estão as respostas } j \text{ e } j'\}}$$

Voltando agora à múltipla $I_1 \times I_2 \times \dots \times I_m = M$ pode-se tirar a partir dela uma distribuição de probabilidade f_M de forma semelhante à descrita acima. Basta dividir o conteúdo de cada cela (i_1, i_2, \dots, i_m) pelo número total de observações.

Pode-se agora passar para uma tabela bidimensional simétrica tomando-se nas margens a união dos fatores do produto $M, \bigcup_{K=1}^m I_K$, e definindo-se cada entrada (i, i') , por analogia à transformação anterior, como a probabilidade do conjunto de elementos de M onde figuram i e i' , multiplicada pelo número total de observações na tabela múltipla.

A interpretação de uma análise efetuada sobre tal representação deve ser feita com cuidado, e por ser algo demasiado específico, por ora não nos aprofundaremos mais no assunto.

1.5 A análise estatística clássica das tabelas de contingência

O número de testes e estimadores estatísticos que foram criados a partir de uma tabela de contingência é bastante considerável. Sociólogos e psicólogos são bastante familiarizados com o parâmetro λ , os índices \mathcal{G} de Goodman e o de Kruskal, e os coeficientes gama e de Kendal, só para citar alguns exemplos. Leo Goodman [9] desenvolveu recentemente uma quantidade enorme de estimadores e técnicas especiais de análise para cruzamentos a ponto de justificar a criação de um pequeno "package" para perfazê-las [10]. No entanto, o estimador mais clássico e conhecido até hoje, base de outros mais sofisticados, é o célebre valor qui-quadrado de Pearson. Devido à sua extrema importância para a compreensão da idéia do método de análise de correspondências apresentamos aqui uma breve, e bastante clássica, discussão de sua construção e propriedades.

1.5.1 Notação e estimadores básicos

Considerando neste parágrafo n_{ij} como números inteiros, definamos os valores $n_i = \sum_{j \in J} n_{ij}$, n.º de observações na linha i ;

$$n_j = \sum_{i \in I} n_{ij}, \text{ n.º de observações na coluna } j;$$

$$n = \sum_{i,j} n_{ij}, \text{ n.º total de observações.}$$

A partir daí, temos naturalmente os seguintes estimadores de máxima verossimilhança:

$$f_{ij} = \frac{n_{ij}}{n}, \text{ estimador de } P[(X_1, X_2) = (i, j)];$$

$f_i = \frac{n_{i.}}{n}$, estimador de $P[X_1 = i]$;

$f_j = \frac{n_{.j}}{n}$, estimador de $P[X_2 = j]$, que dão origem às seguintes medidas de probabilidade:

$$f_{IJ} = \{f_{ij}\}_{(i,j) \in I \times J}, \quad f_I = \{f_i\}_{i \in I}, \quad f_J = \{f_j\}_{j \in J};$$

$$f_{I \times J} = \{f_i \times f_j\}_{(i,j) \in I \times J}$$

1.5.2 O valor qui-quadrado

Foi este estimador criado pelo grande estatístico inglês Karl Pearson por volta de 1900 [18]. Serve ele para solucionar a seguinte questão.

A lei bidimensional do vetor (X_1, X_2) é uma lei produto?

Isto é, $\forall (i, j), P[X_1, X_2 = (i, j)] = P[X_1 = i] \cdot P[X_2 = j]$?

Ora f_{IJ} e $f_{I \times J}$, como definidos em 1.5.1, são respectivamente estimadores da lei de (X_1, X_2) e de uma lei produto em $I \times J$ cujas marginais são iguais às de f_{IJ} . Se a lei real for produto é razoável que f_{IJ} e $f_{I \times J}$ sejam “próximos”, a estatística proposta por Pearson para medir essa proximidade foi:

$$\sum_{i,j} \frac{(n_{ij} - n_{..} \times f_i \times f_j)^2}{n_{..} \times f_i \times f_j} = n_{..} \sum_{i,j} \frac{(f_{ij} - f_i \cdot f_j)^2}{f_i \cdot f_j}$$

Provou ele que quando o tamanho da amostra tende a infinito essa estatística converge fracamente a uma distribuição qui-quadrado com $(|I|-1) \times (|J|-1)$ graus de liberdade. Tal resultado, hoje histórico, permite devido à sua natureza não paramétrica a obtenção de um teste de hipóteses para a pergunta formulada acima (desde que se use amostra “suficientemente” grandes): o muito conhecido teste do qui-quadrado. Observe-se então que para um tamanho de amostra $n_{..}$ fixado, o valor de

$$\sum_{i,j} \frac{(f_{ij} - f_i \cdot f_j)^2}{f_i \cdot f_j}$$

é o responsável pela aceitação ou não da hipótese de independência. Ou seja, esse número, associado a cada possível tabela $I \times J$ provinda de uma amostragem de tamanho $n_{..}$ é uma medida, em suma, do grau de independência no sentido estatístico, entre a classificação das linhas e das colunas. Note-se que mesmo em tabelas de contingência cujas entradas não sejam necessariamente, ou não tenham uma interpretação imediata, como contagens o que foi dito acima continua válido. Ou seja o qui-quadrado de uma tabela é uma medida da riqueza dos condicionamentos recíprocos entre as duas margens do cruzamento.

Com essas idéias em mente, passemos a uma exposição do método.

Capítulo 2

O MÉTODO DE ANÁLISE DE CORRESPONDÊNCIAS

Expõe-se agora o método de análise de correspondências de forma bastante intuitiva, tentando aproximá-lo o mais possível dos outros métodos conhecidos em análise multivariada.

2.1 Nuvens de pontos e eixos fatoriais

Suponhamos, mantendo-se a notação do capítulo anterior, que a hipótese de independência *foi rejeitada* por um teste de qui-quadrado executado em uma tabela $I \times J$. Isso significa, como já se disse, que a lei empírica extraída da tabela é suficientemente distante de uma lei produto. Tomemos então, por exemplo, o conjunto I das linhas e façamos a seguinte associação

$$i \in I \rightarrow f_J^i$$

que tanto pode ser encarada como uma distribuição de probabilidade sobre J ou um vetor em $R^{|J|}$ ou mera função real, de domínio J , cujos valores são

$$\forall j \in J, f_J^i(j) = f_j^i = \frac{f_{ij}}{f_i}$$

Estamos pois, associando a cada i o estimador de máxima verossimilhança da distribuição condicional induzida em J pela fixação do i determinado. Note-se que se a lei f_{iJ} fosse próxima de uma lei independente

$$\forall j \in J, f_j^i \cong \frac{f_i \cdot f_{.j}}{f} = f_j,$$

o que implica na quase igualdade dos $|I|$ vetores f_J^i

Fazendo-se o mesmo com relação ao conjunto J , tomemos então associadas à tabela $I \times J$, e de certa forma a representando, duas configurações ou “nuvens” de pontos, a saber:

$$\begin{aligned} \{f_J^i\}_{i \in I}, & \quad |I| \text{ vetores em } R^{|J|} \\ \{f_I^j\}_{j \in J}, & \quad |J| \text{ vetores em } R^{|I|} \end{aligned}$$

Propõe-se então, reduzir ao máximo a dimensão desses vetores em ambas configurações, preservando no entanto, *da melhor maneira possível, o seu aspecto original*. Tal fato, nos permitirá pela análise da projeção das nuvens em espaços de dimensão inferior examinar as interligações existentes entre os conjuntos I e J , bem como a estrutura induzida em um deles pela classificação do outro

2 1.1 A idéia clássica

Reduzir a dimensão das “observações” é um procedimento típico em análise multivariada, v.g. a análise fatorial e a em componentes principais, remontando a primeira solução novamente a Pearson [17], por volta de 1901. As várias soluções hoje existentes diferem entre si conforme se interprete o que seja “preservar a configuração inicial da melhor maneira possível”. Na maioria dos casos, a idéia básica toma como partida o fato de que reduzir a dimensão dos vetores é *projetá-los* em um subespaço de dimensão inferior. Usa-se então como critério para individualizar tal subespaço o de que a soma das *distâncias* ao quadrado dos pontos iniciais ao subespaço seja mínima. Nessas condições, se χ denota a matriz formada pela colocação em linha dos pontos/vetores centrados, o melhor subespaço de dimensão k é a variedade linear afim que passa pelo centro de gravidade ou ponto médio da nuvem e é paralela ao subespaço dos k autovetores relativos aos k maiores autovalores da matriz $\chi^t \chi$. Esta matriz é muitas vezes chamada matriz de dispersão (Rao [19]), uma vez que ela nos dá a “dispersão” dos dados segundo a norma usada para gerar as distâncias

2.1.2 Generalização

Quando os pontos são afetados de massas quaisquer (que representaremos em geral por uma matriz diagonal M) e a distância é dada por uma forma quadrática definida positiva D , a solução do problema como colocado em 2 1 1 vem da diagonalização de:

$$\zeta = D^{-1} (\chi D)^t M (\chi D)$$

que é a nova matriz de dispersão. Problema semelhante se encontra em mecânica clássica, tendo então os eixos extraídos da matriz o significado de eixos principais de inércia da distribuição de massas.

Note-se que a matriz apresentada em 2.1.1 passa a ser um caso particular da colocada acima com $D =$ matriz identidade e $M = 1/n \times \times$ (matriz identidade) o que faz com que

$$D^{-1} (\chi D)^t M (\chi D) = 1/n \chi^t \chi$$

2.1.3 Em análise de correspondências, a redução da dimensão da configuração é obtida utilizando exatamente as idéias acima, sendo M e D a saber.

- (i) cada ponto f_i^j (resp. f_j^i) será afetado de um peso f_i (resp. f_j), isto implica em se levar mais em conta os vetores que correspondem a i 's (resp. j 's) mais frequentes. O centro de gravidade da configuração, ou esperança matemática desses vetores, é então:

$$E f_j^I = \sum_i f_i \cdot f_j^I = f_j$$

$$(resp. E f_I^j = \sum_j f_j \cdot f_I^j = f_I)$$

(ii) a distância, ao invés de ser gerada pela forma quadrática usual (euclidiana), passa a ser gerada pela seguinte forma quadrática (considerando-se, por exemplo, a nuvem em $R^{|J|}$):

$$(u, v) \rightarrow u^t S^{-1} v \quad u, v \in R^{|J|}$$

onde S matriz $|J| \times |J|$, $s_{jj'} = \delta_{jj'} f_j$

Desse modo, a distância ao quadrado entre $u = (u_1, u_2, \dots, u_{|J|})^t$ e $v = (v_1, v_2, \dots, v_{|J|})^t$ é

$$\sum_j \frac{(u_j - v_j)^2}{f_j}$$

Tal distância, foi chamada do qui-quadrado de centro f_j e o porque de sua introdução ficará claro nas linhas a seguir.

Assim sendo, considerando daqui em diante $R^{|J|}$, temos:

$$S, s_{jj'} = \delta_{jj'} f_j \quad e \quad D = S^{-1} D^t = (S^{-1})^t = S^{-1},$$

$$M, m_{ii'} = \delta_{ii'} f_i,$$

$$\chi, \chi_{ij} = \frac{f_{ij}}{f_i} - f_j;$$

e finalmente $\xi = (S^{-1})^{-1} (\chi S^{-1})^t M (\chi S^{-1}) = (S S^{-1}) \chi^t M \chi S^{-1} = \chi^t M \chi S^{-1}$

$$\text{donde } \xi_{jj'} = \sum_i \frac{(f_{ij} - f_i f_j) (f_{i j'} - f_i f_{j'})}{f_i f_j}$$

será a matriz que se vai diagonalizar.

Mostra-se (ver [2], [3]) que ζ tem o mesmo espectro que uma matriz simétrica especial, sendo seus autovalores compreendidos no intervalo $[0, 1]$. Na realidade, o maior deles é sempre igual a 1 e o autovetor correspondente liga a origem ao centro de gravidade da nuvem.

Ordenemos então os autovalores não nulos de ζ (com exceção do primeiro, igual a 1) em ordem decrescente:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_\alpha \geq \dots \geq \lambda_{|A|}, \quad A \text{ é o conjunto dos índices } \alpha$$

$$tq \lambda_\alpha \neq 0$$

e sejam $u_{1J}, u_{2J}, \dots, u_{|A|J}$ os vetores unitários $((u_{\alpha J})^t S^{-1} u_{\alpha J} = 1)$ correspondentes. Ao conjunto $\{u_{\alpha J}\}_{\alpha \in A}$ chamaremos eixos fatoriais, sua interpretação é, como já disse, igual à da mecânica clássica: colocando-se tais eixos no centro de gravidade da nuvem eles são os seus eixos principais de inércia e a escolha de um determinado número deles individualiza um subespaço ótimo onde se projeta a configuração.

2.2 Fatores e Fórmulas

Vamos agora para cada eixo de inércia $u_{\alpha J}$ definir a função $F_\alpha: I \rightarrow \mathbb{R}$ que a cada ponto i da configuração associa o valor de sua projeção ortogonal sobre $u_{\alpha J}$:

$$\begin{aligned} F_\alpha(i) &= \langle f_J^i - f_J, u_{\alpha J} \rangle \chi^2 = \langle f_J^i, u_{\alpha J} \rangle \chi^2 - \langle f_J, u_{\alpha J} \rangle \chi^2 \\ &= \langle f_J^i, u_{\alpha J} \rangle \chi^2 = (u_{\alpha J})^t S^{-1} f_J^i \end{aligned}$$

uma vez que $u_{\alpha J}$ e f_J são ortogonais.

Obviamente, como temos um sistema de eixos ortogonais, com o conjunto das projeções pode-se reconstruir o ponto:

$$f_J^i = f_J + \sum_{\alpha \in A} F_\alpha(i) u_{\alpha J}$$

logo tendo em conta a definição de f_J^i :

$$\begin{aligned} \forall j \in J, f_{ij} &= f_i \cdot f_j + \sum_{\alpha \in A} F_\alpha(i) f_i \cdot u_{\alpha j} \\ &= f_i \cdot f_j \left(1 + \sum_{\alpha \in A} F_\alpha(i) \frac{u_{\alpha j}}{f_j} \right) \end{aligned}$$

o que nos dá uma *fórmula de reconstrução da tabela inteira a partir das marginais e das projeções nos eixos*; se se projeta então a nuvem em um subespaço de dimensão menor, digamos o plano dos 2 primeiros autovalores, podemos reconstruir a tabela aproximadamente por:

$$\forall i, j \quad f_{ij} \cong f_i \cdot f_j \left(1 + F_1(i) \frac{u_{1j}}{f_j} + F_2(i) \frac{u_{2j}}{f_j} \right)$$

2.2.1 Outras Fórmulas

Multiplicando-se a equação de reconstrução do ponto por f_i e somando-se em i temos:

$$\begin{aligned} \sum_i f_J^i f_i &= f_J \sum_i f_i + \sum_\alpha \sum_i F_\alpha(i) f_i \cdot u_{\alpha J} \\ 0 &= \sum_\alpha \left(\sum_i F_\alpha(i) f_i \right) u_{\alpha J} \end{aligned}$$

e como o conjunto $\{u_{\alpha J}\}_{\alpha \in A}$ é linearmente independente:

$$\forall \alpha \in A, \quad \sum_i F_\alpha(i) f_i = 0$$

o que mostra que o centro de gravidade da nuvem se preserva na projeção nos eixos de inércia, coisa que era de se esperar. É fácil mostrar também que o momento de inércia da configuração em relação a eixo que passa por $u_{\alpha J}$ é igual a λ_{α} , ou seja:

$$\forall \alpha \in A, \sum_i F_{\alpha}^2(i) f_i = \lambda_{\alpha}$$

As duas equações acima escritas também podem ser interpretadas como dando o valor da esperança ($E(F_{\alpha}) = 0$) e da variância ($var F_{\alpha} = \lambda_{\alpha}$) da variável aleatória F_{α} .

2.2.2 As Duas Nuvens

Se houvéssemos considerado a outra nuvem dos $|J|$ pontos em A teríamos é claro uma matriz de dispersão distinta inclusive no tocante à dimensão (a menos que $|J|=|I|$). No entanto, e justamente esta uma das facilidades da métrica do qui-quadrado, mostra-se ([2], [15]) que as matrizes de dispersão das duas nuvens possuem o mesmo conjunto de autovalores não nulos, inclusive em multiplicidade. Definido-se então para cada eixo fatorial $v_{\alpha I}$ a função $G_{\alpha} J \rightarrow IR$ como sendo

$$\begin{aligned} G_{\alpha}(j) &= \langle f_I^j - f_I, v_{\alpha I} \rangle \chi^2 = \langle f_I^j, v_{\alpha I} \rangle \chi^2 = \\ &= (v_{\alpha I})^t W^{-1} f_I^j \end{aligned}$$

onde W é $|I| \times |I|$ e $W_{ij} = \delta_{ij} f_i$.

teremos exatamente o análogo das fórmulas em 2.2.1:

$$\forall \alpha \in A, E G_{\alpha} = \sum_j G_{\alpha}(j) f_j = 0$$

$$\forall \alpha \in A, var G_{\alpha} = E (G_{\alpha})^2 = \sum_j G_{\alpha}^2(j) f_j = \lambda_{\alpha}$$

Como então $\forall \alpha, G_{\alpha}$ e F_{α} são funções centradas e de mesma variância, podemos, dada uma dimensão qualquer para um subespaço fatorial *superpor as projeções de ambas as nuvens nos seus respectivos subespaços obtendo uma representação simultânea de ambas*, ou seja, toda a tabela.

2.2.3 Mais Fórmulas e Princípios

Explorando ainda o inter-relacionamento entre “nuvem das linhas” e “nuvem das colunas” mostra-se nas obras citadas que:

$$\forall \alpha \in A, F_{\alpha} = \lambda_{\alpha}^{\frac{1}{2}} \frac{u_{\alpha I}}{f_I} \text{ e } G_{\alpha} = \lambda_{\alpha}^{\frac{1}{2}} \frac{u_{\alpha J}}{f_J}$$

fórmulas que nos permitem passar dos autovetores de uma matriz para as projeções no autovetor correspondente da outra. Tendo-se em conta a definição das diversas funções envolvidas, podemos escrever a seguinte igualdade pontual tomando-se, por exemplo, a definição de F_α e a igualdade da direita:

$$\begin{aligned} \forall i \in I, F_\alpha(i) &= \langle f_J^i - f_J, u_{\alpha J} \rangle \chi^\alpha = \sum_j f_j^i \times \frac{1}{f_j} \times u_{\alpha j} = \\ &= \sum_j f_j^i \times G_\alpha(j) \lambda_\alpha^{-1/2} \\ \text{logo: } \lambda_\alpha^{-1/2} F_\alpha(i) &= \sum_j G_\alpha(j) f_j^i \end{aligned}$$

A fórmula acima, devido à sua importância denomina-se o *princípio de centro de gravidade* e nos diz que: a abscissa do i -ésimo ponto no eixo fatorial $u_{\alpha J}$ é, a menos do fator $\lambda_\alpha^{1/2}$, o centro de gravidade das abscissas dos pontos da nuvem oposta afetados do vetor de massas f_j^i . Uma vez que o centro de gravidade é um invariante projetivo esse princípio é de grande utilidade na interpretação dos diagramas planos das duas nuvens superpostas.

Finalmente a perfeita simetria entre linhas e colunas pode ser evidenciada mais ainda reescrevendo-se a fórmula de reconstrução da tabela, apresentada no início de 2.2, como:

$$\begin{aligned} f_{ij} &= f_i f_j \left(1 + \sum_{\alpha \in A} F_\alpha(i) \frac{u_{\alpha j}}{f_j} \right) = \\ &= f_i f_j \left(1 + \sum_{\alpha \in A} \lambda_\alpha^{-1/2} F_\alpha(i) G_\alpha(j) \right) \end{aligned}$$

a quantidade $1 + \sum_{\alpha \in A} \lambda_\alpha^{-1/2} F_\alpha(i) G_\alpha(j)$ é pois a expressão da derivada de Radon-Nykodin da medida f_{IJ} com relação a medida $f_{I \times J}$, e linhas e colunas representam um papel formalmente idêntico.

2.2.4 Elementos Suplementares

É comum em muitos casos o desejo de se efetuar uma análise não levando em conta o peso de certas colunas ou linhas. Isto significa aplicar o algoritmo de correspondências para o conjunto das linhas ou colunas com massa não nula e após, projetar os pontos nos planos fatoriais obtidos. Chama-se então a esses pontos de elementos suplementares.

2.3 A Métrica do Qui-quadrado

Conforme se viu nos itens anteriores, o uso da métrica do qui-quadrado permite a grande flexibilidade do método, traduzida no conjunto

de fórmulas e princípios básicos discutidos, e cuja aplicação esperada fique ilustrada com a leitura dos exemplos da Parte II. Gostaríamos de ressaltar que a característica fundamental, responsável pelo interesse e força dos estudos com análise de correspondências, é a de se pode analisar em um mesmo plano fatorial as projeções das duas nuvens de pontos obtendo assim as influências mútuas, as atrações e repulsão entre as variáveis das duas margens da tabela: uma representação pictórica do condicionamento de um grupo pelo outro; fato que gerou a soma o valor de qui-quadrado elevado. No entanto, duas outras características proporcionadas pela métrica em questão são dignas ainda de um olhar mais cuidadoso.

2.3.1 Recuperação do valor χ^2

Como o traço de uma matriz é invariante por diagonalização, podemos, retomando a ξ de 2.1.3, afirmar:

$$\begin{aligned} 1 + \sum_{\alpha \in A} \lambda_{\alpha} &= \text{tr} \xi = \sum_j \xi_{jj} = \\ &= \sum_j \sum_i \frac{(f_{ij} - f_{i \cdot} f_{\cdot j}) (f_{ij} - f_{i \cdot} f_{\cdot j})}{f_{i \cdot} f_{\cdot j}} = \\ &= \sum_{i,j} \frac{(f_{ij} - f_{i \cdot} f_{\cdot j})^2}{f_{i \cdot} f_{\cdot j}} \end{aligned}$$

O último membro dessa igualdade nada mais é que o valor do qui-quadrado da tabela inicial. O que está dito acima é pois que qualquer que seja a dimensão do subespaço que se tome para projetar a nuvem estaremos preservando ao máximo o valor de qui-quadrado da tabela responsável último pela riqueza da configuração.

2.3.2 O princípio da equivalência distribucional

Suponhamos que no conjunto dos $|I|$ pontos em $R^{|J|}$ por exemplo, dois deles sejam muito próximos, a ponto de podermos considerar suas coordenadas como iguais, isto é:

$$\exists i, i' \in I, f_{\cdot j}^i = f_{\cdot j}^{i'}$$

Isto pode significar que ao categorizarmos o conjunto I , as categorias i e i' resultaram *indistinguíveis segundo J* , não havendo por interesse em conservá-las separadas na análise. A métrica do qui-quadrado nos permite passar de I para um conjunto $I_0 = I - \{i, i'\} + \{i_0\}$ de tal forma que as distâncias entre os $|J|$ pontos da outra configuração agora em $IR^{|I|-1}$, permaneçam inalteradas.

Basta para isso definir:

$\forall j \in J, n_{ioj} = n_{ij} + n_{ij'}$, o que implicará a igualdade $\forall j, j' \in J$,

$$\frac{(f_i^j - f_i^{j'})^2}{f_i} + \frac{(f_{i'}^j - f_{i'}^{j'})^2}{f_{i'}} = \frac{(f_{io}^j - f_{io}^{j'})^2}{f_{io}}$$

suficiente para preservar a distância entre qualquer par (j, j')

Assim, a análise de correspondência da configuração em $R^{|J|}$ é inteiramente idêntica à da anterior em $R^{|I|}$, coisa que devia acontecer pois a separação i, i' em nada interferia em J . A esse fato de extrema relevância chama-se *princípio de equivalência distribucional*.

2.4 Considerações Finais

Que ninguém se impressione com o volume de fórmulas apresentado. São elas as justificativas básicas da maioria das interpretações extraídas do resultado de uma correspondência, e a perfeita assimilação de seu significado é fruto de uma interação entre a leitura e aprofundamento teórico do método e o treino adquirido ao se interpretar exaustivamente as saídas de exemplos concretos. Esperamos que o conjunto exposto consista em um ponto de partido sólido para a mestria da técnica. Algumas extensões às idéias aqui lançadas são introduzidas no Capítulo 4. A seguir apresentamos alguns exemplos de caráter geral para ilustrar o exposto até aqui.

Capítulo 3

ALGUNS EXEMPLOS ILUSTRATIVOS

A fim de sedimentar o de certa forma extenso formulário apresentado resolve-se agora passo a passo alguns exercícios. O primeiro tem um caráter um pouco algébrico e visa ilustrar as diversas fases do algoritmo bem como dar uma idéia de uma saída da análise. Os outros revelam certas equivalências interessantes entre as correspondências obtidas das transformações apresentadas no primeiro capítulo.

3.1 Suponhamos uma tabela de descrição lógica a cinco indivíduos i_1, i_2, i_3, i_4, i_5 contra 3 propriedades P_1, P_2, P_3 conforme mostrada abaixo

	P_1	P_2	P_3	marginais linhas
	0	1	0	$1/7$
	0	0	1	$1/7$
	0	1	1	$2/7$
	1	0	0	$1/7$
	1	0	1	$2/7$
marginais colunas	$2/7$	$2/7$	$3/7$	

e vamos submetê-la a uma análise de correspondências.

3 1.1 Para calcularmos os eixos de inércia temos duas opções cor relação à matriz de dispersão a ser diagonalizada. Vamos escolher a relativa à nuvem dos cinco pontos/indivíduos em R^3 , uma vez que uma matriz de dimensão menor e portanto, mais fácil de diagonalizar. Esta matriz, conforme exposto em 2.1.2 tem como entrada geral:

$$\xi_{jj'} = \sum_i \frac{(f_{ij} - f_{i.} f_{.j}) (f_{ij'} - f_{i.} f_{.j'})}{f_{i.} f_{.j}}$$

onde J é o conjunto das propriedades e I o dos indivíduos.

No entanto, no parágrafo citado, foi dito que tal matriz possui o mesmos autovalores que uma certa matriz simétrica, cuja entrada geral apresentamos agora [15]:

$$\sigma_{jj'} = \sum_i \frac{1}{f_{i.}} \frac{f_{ij} f_{ij'}}{f_{.j} f_{.j'}} = \sum_i \frac{1}{n_{i.}} \frac{n_{ij} n_{ij'}}{n_{.j} n_{.j'}}$$

Com um pouco de atenção na fórmula, podemos ver que essa matriz a entrada $\sigma_{jj'}$ é o produto de 3 outras, a saber:

1. Uma matriz $J \times I$ igual à transposta da tabela dada, dividir do-se ainda cada cela pela raiz quadrada do total da coluna (j) correspondente;

2. Uma matriz diagonal $I \times I$ cujas entradas são os inversos de totais das linhas (i) respectivas;

3. A transposta da matriz descrita em 1.
No nosso caso concreto temos então:

$$\begin{bmatrix} 0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 & 0 \\ 0 & 1/\sqrt{3} & 1/\sqrt{3} & 0 & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 0 & 1/\sqrt{2} & 0 \\ 0 & 0 & 1/\sqrt{3} \\ 0 & 1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{3} \end{bmatrix} =$$

$$= \begin{bmatrix} 1/2 & 0 & \frac{\sqrt{6}}{12} \\ 0 & 3/4 & \frac{\sqrt{6}}{12} \\ \frac{\sqrt{6}}{12} & \frac{\sqrt{6}}{12} & 2/3 \end{bmatrix}$$

simétrica como era de se esperar.

Diagonalizando-se esta matriz pelo processo algébrico clássico, vamos encontrar os autovalores:

$$\lambda_0 = 1$$

$$\lambda_1 = \frac{9}{12}$$

$$\lambda_2 = \frac{5}{12}$$

3.1.2 Extraindo agora os autovetores unitários, segundo a norma euclidiana, pelo processo usual, temos:

$$\lambda_0 \cdot X_{0J} = \left(\frac{\sqrt{2}}{\sqrt{7}}, \frac{\sqrt{2}}{\sqrt{7}}, \frac{\sqrt{3}}{\sqrt{7}} \right)$$

$$\lambda_1 \cdot X_{1J} = \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 0 \right)$$

$$\lambda_2 \cdot X_{2J} = \left(\frac{-\sqrt{3}}{\sqrt{14}}, \frac{\sqrt{3}}{\sqrt{14}}, \frac{2\sqrt{7}}{7} \right)$$

Porém atenção, esses 3 vetores que formam uma base ortonormal do R^3 segundo a métrica de Euclides, são relativos à matriz que obtivemos de espectro equivalente à de dispersão.

Para obter os autovetores desta, temos que aplicar a transformação linear:

$$S^{1/2} = \begin{bmatrix} \frac{\sqrt{2}}{\sqrt{7}} & 0 & 0 \\ 0 & \frac{\sqrt{2}}{\sqrt{7}} & 0 \\ 0 & 0 & \frac{\sqrt{3}}{\sqrt{7}} \end{bmatrix}$$

Em geral, a transformação $S^{1/2}$ definida como,

$$(S^{1/2})_{jj'} = \delta_{jj'} (f_j)^{1/2}$$

levará os autovetores unitários Euclides da matriz auxiliar, em autovetores unitários qui-quadrado, da matriz de dispersão ([2], [15]). Assim teremos:

$$X_{0J} \rightarrow S^{1/2} X_{0J} = u_{0J} = \left(\frac{2}{\gamma}, \frac{2}{\gamma}, \frac{3}{\gamma} \right)$$

$$X_{1J} \rightarrow S^{1/2} X_{1J} = u_{1J} = \left(\frac{\sqrt{7}}{\gamma}, -\frac{\sqrt{7}}{\gamma}, 0 \right)$$

$$X_{2J} \rightarrow S^{1/2} X_{2J} = u_{2J} = \left(-\frac{\sqrt{3}}{\gamma}, -\frac{\sqrt{3}}{\gamma}, \frac{2\sqrt{3}}{\gamma} \right)$$

Convidamos o leitor a verificar que $\{u_{0J}, u_{1J}, u_{2J}\}$ são uma base ortonormal do R^3 segundo a métrica do qui-quadrado de centro

$$f_J = \left(\frac{2}{\gamma}, \frac{2}{\gamma}, \frac{3}{\gamma} \right)$$

3 1 3 Lembremos que os pontos de nossa nuvem são as distribuições condicionais sobre as colunas geradas pela fixação de cada linha, ou seja:

$$i_1 = (0, 1, 0)$$

$$i_2 = (0, 0, 1)$$

$$i_3 = \left(0, \frac{1}{2}, \frac{1}{2} \right)$$

$$i_4 = (1, 0, 0)$$

$$i_5 = \left(\frac{1}{2}, 0, \frac{1}{2} \right)$$

ora, o centro de gravidade da distribuição de massas $(1/\gamma, 1/\gamma, 2/\gamma, 1/\gamma, 1/\gamma)$ colocada nesses pontos é:

$$\begin{aligned} \frac{1}{\gamma} (0, 1, 0) + \frac{1}{\gamma} (0, 0, 1) + \frac{2}{\gamma} \left(0, \frac{1}{2}, \frac{1}{2} \right) + \frac{1}{\gamma} (1, 0, 0) + \frac{2}{\gamma} \left(\frac{1}{2}, 0, \frac{1}{2} \right) &= \\ &= \left(\frac{2}{\gamma}, \frac{2}{\gamma}, \frac{3}{\gamma} \right) \end{aligned}$$

que é exatamente o vetor u_{0J} , pois como se viu em 2.1.2 o autovetor relativo ao maior autovalor liga a origem dos eixos coordenados ao centro de gravidade da nuvem, ponto de origem dos eixos de inércia.

Como o primeiro autovetor não nos interessa, vamos obter as projeções da nuvem nos outros dois restantes, gerando as funções F_1 e F_2 ; assim temos.

$$F_1(i_1) = \langle u_1, i_{1J} \rangle_{\chi^2} = \frac{-\sqrt{7}}{\gamma} \times \frac{\gamma}{2} = -\frac{\sqrt{7}}{2}$$

$$F_1(i_2) = \langle u_2, i_{1J} \rangle_{\chi^2} = 0$$

$$F_1(i_3) = \langle u_3, i_{1J} \rangle_{\chi^2} = \frac{1}{2} \times \frac{-\sqrt{7}}{\gamma} \times \frac{\gamma}{2} = -\frac{\sqrt{7}}{4}$$

$$F_1(i_4) = \langle u_4, i_{1J} \rangle_{\chi^2} = \frac{\sqrt{7}}{4} \times \frac{\gamma}{2} = \frac{\sqrt{7}}{2}$$

$$F_1(i_5) = \langle u_5, i_{1J} \rangle_{\chi^2} = \frac{\sqrt{7}}{\gamma} \times \frac{1}{2} \times \frac{\gamma}{2} = \frac{\sqrt{7}}{4}$$

Calculando a esperança e a variância de F_1 temos:

$$E F_1 = -\frac{\sqrt{\gamma}}{2} \times \frac{1}{\gamma} + 0 \times \frac{1}{\gamma} + \frac{\sqrt{\gamma}}{4} \times \frac{2}{\gamma} + \frac{\sqrt{\gamma}}{2} \times \frac{1}{\gamma} + \frac{\sqrt{\gamma}}{4} \times \frac{2}{\gamma} = 0$$

$$\text{var } F_1 = \left(-\frac{\sqrt{\gamma}}{2}\right)^2 \frac{1}{\gamma} + 0^2 \times \frac{1}{\gamma} + \left(\frac{-\sqrt{\gamma}}{4}\right)^2 \frac{2}{\gamma} + \left(\frac{\sqrt{\gamma}}{2}\right)^2 \frac{1}{\gamma} + \left(\frac{\sqrt{\gamma}}{4}\right)^2 \frac{2}{\gamma} =$$

$$= \frac{3}{4} = \frac{\theta}{12} = \lambda_1$$

conforme as fórmulas de 2.2.1.

3.1.4 Tomemos agora, por exemplo, o ponto i_4 , e façamos ainda alguns cálculos com ele (ver Figura 1):

i) contribuição de i_4 à inércia relativa à u_{1J} :

basta ver que a parcela referente a i_4 em $\text{var } F_1$, vale:

$$F_1(i_4) \times f_{i_4} = \frac{\gamma}{4} \times \frac{1}{\gamma} = \frac{1}{4}$$

e logo a contribuição relativa é.

$$\frac{1}{\lambda_1} (F_1(i_4) \times f_{i_4}) = \frac{1}{4} : \frac{3}{4} = 33,3\%$$

ii) ângulo de $i_4 - f_J$ com o eixo u_{1J} :

$$\cos i_4 - f_J, u_{1J} = \frac{\langle i_4 - f_J, u_{1J} \rangle_{\mathcal{X}^2}}{\|i_4 - f_J\|_{\mathcal{X}^2}} = \frac{\langle i_4, u_{1J} \rangle_{\mathcal{X}^2}}{\|i_4 - f_J\|_{\mathcal{X}^2}} =$$

$$= \frac{F_1(i_4)}{\frac{\sqrt{\gamma}}{2}} = \frac{\frac{\gamma}{\gamma^2} \times \frac{\gamma}{2} + \frac{2^2}{\gamma^2} \times \frac{\gamma}{2} + \frac{3^2}{\gamma^2} \times \frac{\gamma}{3}}{\frac{25}{14} + \frac{2}{\gamma} + \frac{\gamma}{7}} =$$

$$= \frac{\sqrt{\gamma}}{2} \times \frac{\sqrt{2} \cdot \sqrt{\gamma}}{\sqrt{\gamma} \cdot \sqrt{5}} = \frac{\sqrt{14}}{2\sqrt{5}}$$

Caso quiséssemos calcular o ângulo do vetor $i_4 - f_J$ com o plano fatorial $u_{1J} \times u_{2J}$, bastaria notar que:

$$\text{proj}_{u_{1J} \times u_{2J}} i_4 = \text{proj}_{u_{1J} \times u_{2J}} i_4 - f_J = F_1(i_4) u_{1J} + F_2(i_4) u_{2J}$$

e logo:

$$\cos i_4 - f_J, u_{1J} \times u_{2J} = \frac{\langle i_4 - f_J, F_1(i_4) u_{1J} + F_2(i_4) u_{2J} \rangle_{\mathcal{X}^2}}{\|i_4 - f_J\|_{\mathcal{X}^2} \|F_1(i_4) u_{1J} + F_2(i_4) u_{2J}\|_{\mathcal{X}^2}}$$

$$= \frac{F_1^2(i_4) + F_2^2(i_4)}{\|i_4 - f_J\|_{\mathcal{X}^2} \sqrt{F_1^2(i_4) + F_2^2(i_4)}} = \frac{\sqrt{F_1^2(i_4) + F_2^2(i_4)}}{\|i_4 - f_J\|_{\mathcal{X}^2}}$$

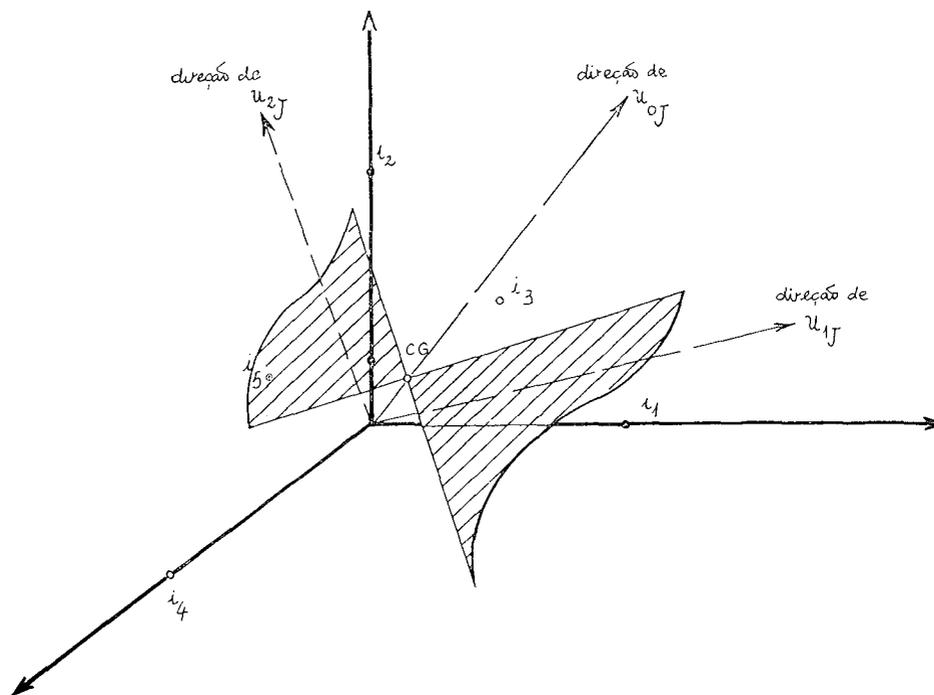
logo, pode-se dizer que:

$$\cos^2 i_4 - f_J, u_{1J} \times u_{2J} = \cos^2 i_4 - f_J, u_{1J} + \cos^2 i_4 - f_J, u_{2J},$$

o que era de se esperar pois $\{u_{1J}, u_{2J}\}$ é uma base ortogonal do plano $u_{1J} \times u_{2J}$.

FIGURA 1

A nuvem dos indivíduos em R^3 , os 3 eixos qui-quadrado e o primeiro plano fatorial passando pelo centro de gravidades da nuvem



3.1.5 Se quiséssemos considerar os pontos/propriedades no espaço R^5 dos indivíduos, teríamos:

$$P1 = \left(0, 0, 0, \frac{1}{2}, \frac{1}{2} \right)$$

$$P2 = \left(\frac{1}{2}, 0, \frac{1}{2}, 0, 0 \right)$$

$$P3 = \left(0, \frac{1}{3}, \frac{1}{3}, 0, \frac{1}{3} \right)$$

Aplicando as fórmulas de transformação de 2.2.3, teríamos rapidamente para o eixo relativo a λ_1 , por exemplo:

$$G_1(P1) = \frac{9}{12} \frac{u_{1P1}}{f_{P1}} = \frac{3}{2\sqrt{3}} \times \frac{\sqrt{7}}{7} \times \frac{7}{2} = \frac{3}{4} \frac{\sqrt{7}}{\sqrt{3}}$$

$$G_1(P2) = \frac{9}{12} \frac{u_{1P2}}{f_{P2}} = \frac{3}{2\sqrt{3}} \times \frac{-\sqrt{7}}{7} \times \frac{7}{2} = -\frac{3}{4} \frac{\sqrt{7}}{\sqrt{3}}$$

$$G_1(P3) = \frac{9}{12} \frac{u_{1P3}}{f_{P3}} = \frac{3}{2\sqrt{3}} \times 0 \times \frac{7}{3} = 0$$

Apenas para ilustrar, poderíamos obter primeiro o autovetor v_{II} usando o mesmo conjunto de fórmulas:

$$v_{II} = \lambda_I^{-1/2} F_I \times f_I = \frac{2\sqrt{3}}{3} \left(\frac{-\sqrt{\gamma}}{2} \times \frac{1}{\gamma}, 0 \times \frac{1}{\gamma}, \frac{-\sqrt{\gamma}}{4} \times \frac{2}{\gamma}, \frac{\sqrt{\gamma}}{2} \times \frac{1}{\gamma}, \frac{\sqrt{\gamma}}{4} \times \frac{2}{\gamma} \right) = \left(\frac{-\sqrt{\gamma}}{\gamma\sqrt{3}}, 0, \frac{-\sqrt{\gamma}}{\gamma\sqrt{3}}, \frac{\sqrt{\gamma}}{\gamma\sqrt{3}}, \frac{\sqrt{\gamma}}{\gamma\sqrt{3}} \right)$$

e daí G_I , por exemplo:

$$G_I(P1) = \langle P1, v_{II} \rangle_{\lambda^e} = 0 + 0 + 0 + \gamma \times \frac{1}{2} \times \frac{\sqrt{\gamma}}{\gamma\sqrt{3}} + \frac{1}{2} \times \frac{\sqrt{\gamma}}{\gamma\sqrt{3}} \times \frac{\gamma}{2} = \frac{3}{4} \sqrt{\frac{\gamma}{3}}$$

o que confere com o resultado obtido

3 1.6 Finalmente apresentamos na Figura 2 o resultado da projeção de ambas as nuvens no plano dos dois primeiros eixos. Note-se como após essas obscurecedoras observações algébricas, a saída gráfica nos proporciona uma rápida interpretação visual da tabela (Os indivíduos/linhas foram circundados por retângulos para diferenciá-los das colunas/propriedades; i_1, i_2, \dots, i_5 são respectivamente $LI1, LI2, \dots, LI5$ e $P1, P2, P3$ são $CO1, CO2, CO3$).

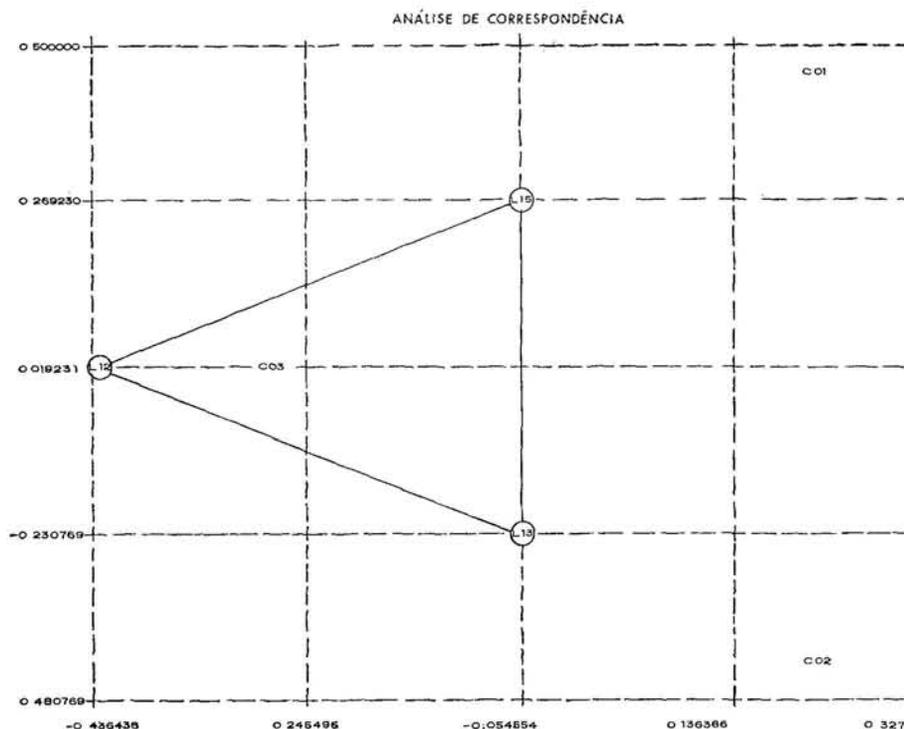
Os indivíduos e as propriedades se distribuem simetricamente em relação à horizontal que passa por i_2 e $P3$. Este ponto, como se vê, está a menos de um fator de escala, no centro de gravidade do triângulo formado por i_2, i_3 e i_5 uma vez que a condicional nas linhas ao se fixar $P3$ vale zero para i_1 e i_4 (veja item 2.2.3). Por sua vez, i_1 e i_4 formam com i_3 e i_5 respectivamente, dois segmentos cujo fulcro está cerca de $P2$ e $P1$ a saber. Em resumo a figura mostra claramente a independência de i_2 das primeiras propriedades e de i_1 e i_4 da última. O gráfico sugere ainda uma permutação na tabela original que a coloca sob a forma mais sugestiva:

	$P2$	$P3$	$P1$
i_1	1	0	0
i_3	1	1	0
i_2	0	1	0
i_5	0	1	1
i_4	0	0	1

Observe-se ainda que o primeiro eixo de inércia (vertical), como acontecer, é o que separa mais os pontos. Assim, por exemplo, o conjunto dos indivíduos encontra-se perfeitamente escalonado por suas projeções nesse eixo. Obviamente tal escala é função do comportamento de i_1, \dots, i_5 frente a $P1, P2$ e $P3$ o que se reflete tanto analogicamente na tabela rearranjada como na figura pela oposição de i_1 e i_2, i_3 e i_4 . Este efeito, extremamente comum em várias situações, denomina-se efeito Gutman e será abordado com mais detalhe no próximo capítulo.

FIGURA 2

A projeção de ambas as nuvens, no primeiro plano fatorial



3.1.7 Para concluir vamos empregar a fórmula de reconstituição (2.2 início) para obter uma aproximação da cela $(i_1, P1)$ usando o meio eixo fatorial:

$$\begin{aligned}
 f_{i_1 P1} &\cong f_{i_1} f_{P1} + \lambda^{1-\frac{1}{2}} F_1(i_1) G_1(P1) = \frac{1}{7} \times \frac{2}{7} + \left(\frac{9}{12}\right)^{-\frac{1}{2}} \times \\
 &\times \frac{\sqrt{7}}{2} \times \frac{3}{4} \frac{\sqrt{7}}{\sqrt{3}} = \frac{2}{49} \times 1 - \frac{7}{4} = -\frac{3}{98} \cong 0,03
 \end{aligned}$$

que aproxima o valor zero. Não deve causar espanto o fato de se ter encontrado um valor negativo, uma vez que se efetuou um truncamento na fórmula que reproduz exatamente $f_{i_1 P1}$.

3.2 (Nota: Apenas para facilitar os cálculos, consideramos nesse exemplo que as entradas estão normalizadas, i. e., $\sum_{i,j} n_{ij} = 1$)

Suponhamos uma tabela de contingência $I \times J$ onde se aplicou a transformação apresentada em 1.4.2, isto é, passa-se das entradas n_{ij} para uma tabela simétrica $J \times J$ a entradas $\sum_i \frac{n_{ij} n_{ij'}}{n_{i.}} t_{jj}$

Conforme já se mencionou, tal transformação é frequentemente usada pelos algoritmos para a obtenção de estruturas hierárquicas via correspondências. Propomo-nos a explorar várias relações entre as duas tabelas.

3.2.1 Um princípio de equivalência distribucional

Admitamos que na tipologia de I existam dois elementos indistinguíveis segundo J , ou seja, para i e $i' \in I$ vale a relação

$$\frac{n_{ij}}{n_{i'j}} = \frac{n_{ij'}}{n_{i'j'}}$$

para todo par j, j' . Nada mais natural do que fazermos a agregação definida em 2.3.2, ou seja, juntá-los numa mesma linha resultante da soma das anteriores. Vamos provar que a nova tabela simetrizada mantém-se inalterada. Notemos primeiro que, usando a notação do parágrafo citado:

$$\frac{n_{ij}}{n_{i'j}} = \frac{n_{ij'}}{n_{i'j'}} \implies \frac{n_{ij}}{n_{ij'}} = \frac{n_{i'j}}{n_{i'j'}} = \frac{n_{ij} + n_{i'j}}{n_{ij'} + n_{i'j'}} = \frac{n_{i_0j}}{n_{i_0j'}}$$

logo, somando em j :

$$\frac{n_{i.}}{n_{i'j'}} = \frac{n_{i'.}}{n_{i'j'}} = \frac{n_{i_0.}}{n_{i_0j'}}$$

Ora, as únicas alterações no somatório que define a nova entrada são relativas às parcelas contendo i e i' ; tendo em vista a igualdade acima, podemos escrever:

$$\begin{aligned} n_{ij} \times \frac{n_{ij'}}{n_{i.}} + n_{i'j} \times \frac{n_{i'j'}}{n_{i'.}} &= n_{ij} \times \frac{n_{i_0j'}}{n_{i_0.}} + n_{i'j} \times \frac{n_{i_0j'}}{n_{i_0.}} = \\ &= \frac{n_{i_0j'}}{n_{i_0.}} (n_{ij} + n_{i'j}) = \frac{n_{i_0j'} n_{i_0j}}{n_{i_0.}} \end{aligned}$$

que é a nova parcela do somatório, estando pois provada a afirmação. Esta invariância era esperada pois na nova tabela procura-se particionar J segundo os atributos I e a partição resultante sendo extraída de uma análise de correspondências dever-se-ia manter invariável com a agregação feita.

3.2 2 Seja agora $\{(u_{\alpha I}, u_{\alpha J}, \lambda_{\alpha})\}_{\alpha \in A}$ o conjunto dos autovetores $u_{\alpha I}, u_{\alpha J}$ com os respectivos autovalores associados λ_{α} , não nulos. Tentemos exprimir os autovetores e autovalores da matriz transformada em função deles. Sejam de início algumas relações entre os parâmetros da transformada e da original:

$$i) \quad t_j = \sum_{j'} t_{j'j} = \sum_{j'} \sum_i \frac{n_{ij'} n_{ij}}{n_i} = \sum_i \frac{n_{ij}}{n_i} \sum_{j'} n_{ij'} = \sum_i n_{ij} = n_j$$

ou seja, as marginais da transformada coincidem com as marginais em J da primeira tabela.

$$ii) \quad t_{j'}^j = \frac{t_{j'j}}{t_j} = \frac{1}{n_j} \sum_i \frac{n_{ij} n_{ij'}}{n_i} = \sum_i \frac{n_{ij} n_{ij'}}{n_i n_j}$$

essa relação liga a condicional da tabela transformada com elementos da tabela inicial.

Seja então v_J um autovetor normalizado χ^2 da tabela $J \times J$ relativo ao autovalor μ . Relembrando o parágrafo 2 2.3 temos:

$$\mu^{1/2} \frac{v_J}{t_j} = G_{\alpha} e G_{\alpha}(j) = \langle t_{j'}^j, v_J \rangle \chi^2$$

ou seja

$$\forall j \in J, \mu^{1/2} \frac{v_j}{t_j} = \sum_{j'} t_{j'}^j \times v_{j'} \times \frac{1}{t_{j'}} \\ \mu^{1/2} \frac{v_j}{t_j} = \sum_{j'} t_{j'}^j \times \frac{v_{j'}}{t_{j'}} ;$$

tendo em vista os itens i e ii acima, podemos escrever:

$$\forall j \in J \mu^{1/2} \frac{v_j}{n_j} = \sum_{j'} \left(\sum_i \frac{n_{ij} n_{ij'}}{n_i n_{j'}} \right) \frac{v_{j'}}{n_{j'}}, \quad (1)$$

recordando-se a discussão no início do exemplo anterior, podemos afirmar que se $x_{\alpha J}$ é um autovetor euclides da correspondência $I \times I$ relativo ao autovalor λ_{α} temos

$$\forall j \in J \lambda_{\alpha} x_{\alpha j} = \sum_{j'} \sum_i \frac{n_{ij} n_{ij'}}{n_i n_{j'}} x_{\alpha j'} \\ \text{ou } \lambda_{\alpha} \frac{x_{\alpha j}}{n_j} = \sum_{j'} \sum_i \frac{n_{ij} n_{ij'}}{n_i n_{j'}} \times \frac{x_{\alpha j'}}{n_{j'}} \quad (2)$$

e como sabemos $\frac{x_{\alpha J}}{n_J}$ é o autovetor qui-quadrado correspondente.

Logo de (1), (2) e propriedades elementares do espectro de uma matriz podemos concluir que:

$\forall \alpha \in A$ a cada autovalor λ_α da correspondência $I \times J$ corresponde um autovalor $\mu_\alpha = \lambda_\alpha^2$ da simetrizada $J \times J$, ambos com o mesmo autovetor.

3.3 Consideremos a simetriação de uma tabela disjuntiva completa $I \times J$. Recordemos (1.2) que o melhor exemplo de uma tal tabela é o de um questionário $Q = \{q\}$ que foi submetido a um conjunto I de indivíduos. A margem $J = \bigcup_{\alpha \in Q} J_\alpha$ é a união em Q de todas as respostas possíveis J_α à cada pergunta α . Como vimos uma disjuntiva completa é comumente simetrizada numa $J \times J$ mediante a transformação

$$s_{j', j''} = \# \{i \in I \mid n_{ij'} = n_{ij''} = 1\},$$

ou seja, cada entrada nos dá o número de pessoas que escolheram as respostas j' e j'' . Propomos, como anteriormente, a relacionar as análises de correspondências extraídas das duas tabelas

3.3.1 Uma vez que cada indivíduo escolhe uma e única resposta para cada pergunta q podemos afirmar:

$\forall i \in I, \sum_j n_{ij} = |Q|$, i. e., é o número de perguntas do questionário. E como a tabela $I \times J$ é uma tabela lógica:

$$S_{j', j''} = \# \{i \in I \mid n_{ij'} = n_{ij''} = 1\} = \sum_{i \in I} n_{ij'} n_{ij''}$$

Mediante uma simples álgebra temos então

$$\begin{aligned} S_{j', j''} &= \sum_{i \in I} n_{ij'} n_{ij''} = \sum_{i \in I} n_{ij'} n_{ij''} \times \frac{\sum_{j \in J} n_{ij}}{\sum_{j \in J} n_{ij}} = \\ &= |Q| \times \sum_{i \in J} \frac{n_{ij'} n_{ij''}}{\sum_{j \in J} n_{ij}} \end{aligned}$$

A expressão acima nos mostra que todas as entradas $S_{j', j''}$ são a menos do fator $|Q|$ idênticas às entradas $t_{j', j''}$ da simetriação discutida no exemplo anterior. Logo vale o mesmo resultado, qual seja, os autovalores da nova matriz são o quadrado dos da matriz anterior, sendo idênticos os autovetores respectivos.

3.3.2 Relembrando a representação obtida no Capítulo 1 de tabelas múltiplas em tabelas bidimensionais disjuntivas completas ou simétricas, o resultado acima consolida tais representações como alternativas para a análise de tabelas múltiplas.

Capítulo 4

ALGUMAS EXTENSÕES E RESULTADOS PARTICULARES

Nesse capítulo apresenta-se algumas correspondências particulares, com o intuito de se dar idéia dos tipos de estudos mais profundos sobre o método. Aqueles com uma formação estatística e matemática mais sólida podem não se saciar com o caráter relativamente geral das discussões e ansiar por maiores detalhes que podem ser encontrados na bibliografia ou, quem sabe, estão à espera do trabalho deles próprios para serem solucionados.

4.1 Efeito Guttman

Muitas vezes pela estrutura mesma de uma dada tabela, podemos prever qual será o resultado da correspondência nos diversos planos fatoriais. De modo contrário, determinados resultados padrões sobre os planos fatoriais podem sugerir ou implicar a existência de estruturas específicas, não percebidas, na tabela. Dentro dessa linha, uma das saídas mais extensamente estudadas é a que recebeu o nome geral de Efeito Guttman.

4.1.1 Definição e exemplos

Diz-se que há efeito Guttman, quando em uma análise de correspondências os diversos fatores extraídos guardam uma relação analítica entre si. Como os fatores são sempre não correlacionados esta relação jamais pode ser linear.

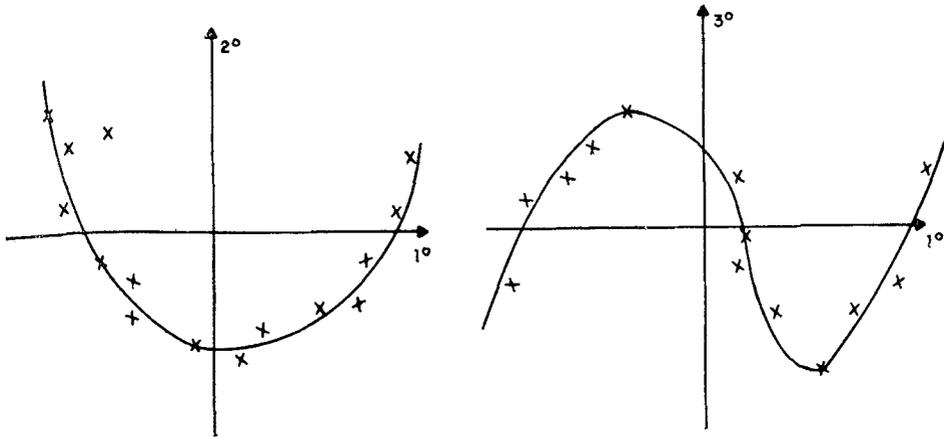
Lembrando ainda que como os fatores são na realidade funções e que portanto, a relação deve valer para cada ponto do domínio, o caso clássico mais comum é o em que o p -ésimo fator é um polinômio de grau p no primeiro. Tal caso produz desenhos típicos nos planos fatoriais, como indicado na Figura 1.

Havendo uma relação como a acima mencionada, pode-se projetando a curva no eixo adequado, extrair naturalmente uma ordenação das variáveis/ponto que representavam uma margem pelas variáveis da outra. Constrói-se assim uma *escala*, ou seja, *uma representação métrica unidimensional* de um conjunto de variáveis que pretendem representar certo fenômeno.

Este tipo de problema foi estudado inicialmente por sociólogos e psicólogos, tendo sido Louis Guttman [2] um dos primeiros a produzir resultados analíticos consistentes para a construção de escalas sociométricas. A tabela de contingência padrão, que ele utilizou para ilustrar esse efeito, é uma tabela desdobrada provinda de um questionário contendo n perguntas (conjunto J) que só admitem respostas sim (1) ou não (0), tendo sido este questionário dado a n pessoas que

FIGURA 1

Os polinômios ortogonais de um efeito Guttman nos dois primeiros planos fatoriais



respondem de forma seguinte: o primeiro responde sim a todas, o segundo sim a todas menos a primeira, e assim por diante. Colocando-se as perguntas e os indivíduos em ordem, a tabela fica com o seguinte aspecto:

	q_1^+	q_2^+		q_n^+	q_1^-	q_2^-	q_n^- (perg. desdobradas)
i_1	1	1		1	0	0	0
i_2	0	1	..	1	1	0	0
peessoas i_3	0	0		1	1	1	0
	..						
i_n	0	0		0	1	1	1

A análise de correspondência dessa matriz nos dá para a nuvem de indivíduos $I = \{1, 2, \dots, n\}$ um primeiro fator da forma $F_1(i) = ai + b$ e um segundo da forma $F_2(i) = a' i^2 + b'i + c'$ havendo portanto efeito Guttman:

$$F_2 = \frac{a'}{a^2} (F_1 - b) + \frac{b'}{a} (F_1 - b) + c'$$

Em geral originando uma tabela um resultado próximo ao da Figura 1, ao se reescrever a tabela segundo a ordenação dada pela escala ela toma uma forma como a mostrada na Figura 2, que nada mais é do que uma aproximação à tabela ideal acima apresentada.

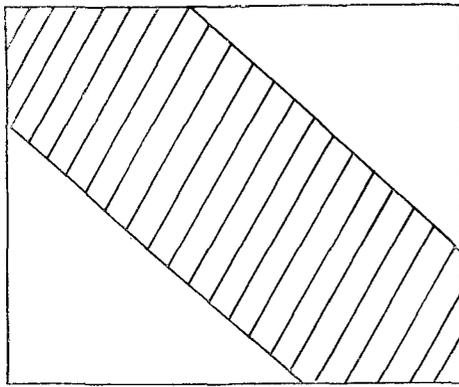


FIGURA 2

Uma tabela de contingência próxima a uma tabela Guttman, após o reagrupamento sugerido pela saída.

(A parte não hachuriada corresponde a zeros).

4 1.2 Passagens ao Limite

Uma vez que a análise de correspondências nada mais é que uma representação de uma distribuição de probabilidades definida sobre um produto cartesiano a dois fatores, nada impede, teoricamente, que se estude como se comporta tal representação quando os fatores do produto passam a ser conjuntos/variáveis aleatórias contínuos. Em muitos casos especiais de efeito Guttman, conclusões advindas do estudo teórico de uma “tabela contínua” podem ser adequadamente ajustadas e aplicadas em tabelas reais. Apenas para dar ao leitor um sentimento de como se processam tais generalizações mostremos a criação de um análogo contínuo à tabela clássica de Guttman descrita em 4.1.1.

O conjunto das pessoas e o das respostas passam a ser indicados pelos parâmetros contínuos $x, u \in (0,1)$ e o das respostas desdobradas por $y \in (0,2)$. Desse modo a resposta afirmativa (resp. negativa) à questão q_u recebe o valor $y=u$ (resp. $y=1+u$). O análogo contínuo da matriz anterior será então associar ao indivíduo x respostas afirmativas às questões $q_u, u > x$ e negativas às que $u < x$. A cada “linha” x da nova “matriz” corresponde a função indicador $I_{(x, x+1)}$ definida sobre o conjunto $(0,2)$. A “tabela” então nada mais será do que uma lei de probabilidade em $X \times Y = (0,1) \times (0,2)$ definida pela seguinte densidade em relação a medida de Lebergue:

$$\forall (x, y) \in X \times Y, \quad p(x, y) = I_{(x, x+1)}(y)$$

ou seja, como a sua correspondente discreta, a nova tabela contínua dá “massa” ao conjunto hachuriado na Figura 3. As densidades marginais serão portanto:

$$P_x(x) = 1, \quad x \in (0, 1)$$

$$P_y(y) = \begin{cases} y, & y \in (0, 1) \\ 2 - y, & y \in (1, 2) \end{cases}$$

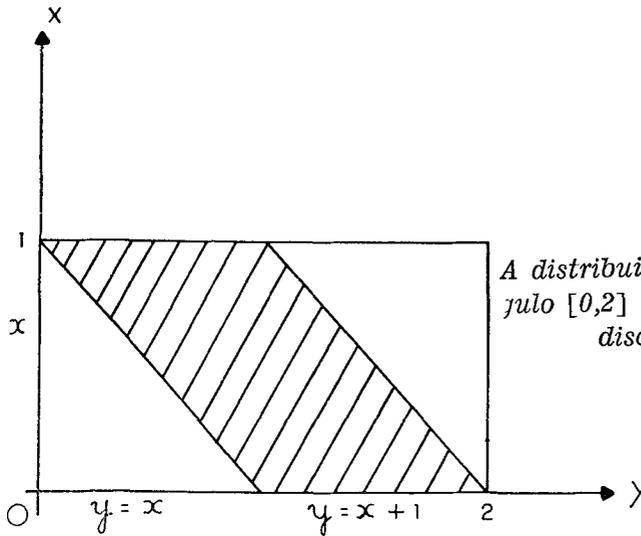


FIGURA 3

A distribuição de massa no retângulo $[0,2] \times [0,1]$ análoga ao caso discreto da Figura 2

Foge ao espírito do presente trabalho aprofundar-se mais a partir daqui, aos interessados urge que consultem o Capítulo 7 da parte B de [2], vol. II Citamos apenas que definindo-se convenientemente operadores projeção contínuos, análogos aos do caso discreto obtém-se da sua diagonalização autovalores da forma:

$$\lambda_i = \frac{1}{i(i+1)}, \quad 1 \leq i \leq p, \quad p \in \mathbb{N}$$

e autovetores respectivos $u_i(x) = (2i+1)^{1/2} P_i(2x-1)$, $x \in (0,1)$, onde P_i é o i -ésimo polinômio de Legendre. A partir daí prova-se que para a “nuvem” pessoas, o segundo e terceiro fatores guardam entre si as relações:

$$F_2 = \left(\frac{5}{24}\right)^{1/2} (2 F_1^2 - 1)$$

$$F_3 = \left(\frac{7}{18}\right)^{1/2} 6^{-1} (10 F_1^3 - 9 F_1)$$

o que mostra que o efeito Guttman foi preservado.

A importância dessas análises contínuas reside pois no fato que dando-se ao modelo contínuo distribuições de probabilidade especiais podemos calcular realmente, com relativa facilidade, quais as “figuras” resultantes na representação. Sabe-se, por exemplo, que se a lei em $X \times Y$ é uma normal bidimensional os autovetores serão a menos de ligeira modificação os polinômios de Hermite e os fatores de uma das margens apresentarão efeito Guttman clássico. Se, por exemplo, uma das margens (X) é homeomorfa ao círculo unitário e a outra (Y) possui uma estrutura onde faça sentido agir o grupo das rotações no cír-

culo, sabe-se que se a lei de probabilidade possui uma densidade invariante por rotações há efeito Guttman em X na forma de um desdobramento de Fourier.

Assim sendo, o aparecimento na prática, nos diversos planos fatoriais de relações específicas como as acima descritas pode ser o indicio que uma das margens possua estrutura semelhante à que implica teoricamente o surgimento de tal relação, e que havia se mantido até ali velada ao pesquisador. Além disso esses estudos fornecem um bom teste teórico da validade e eficácia das propriedades da análise de correspondências.

4.2 Testes sobre os resultados em correspondências

Embora pela sua própria estrutura e filosofia e, muitas vezes, pelo próprio tamanho da amostra envolvida, não faça sentido o preocupar-se com a construção de testes estatísticos clássicos sobre os resultados de uma correspondência isso não invalida nem implica em se deixar sem verificação ou prova os resultados de uma análise conduzida. Assim é que, embora ainda não muito desenvolvidas e estudadas, existem algumas provas, ou técnicas, ou testes que pretendem verificar a validade dos resultados da análise.

4 2 1 Procedimentos usuais

É comum estudar-se a estabilidade de resultados obtidos modificando-se a tabela de contingência de forma a refletir a variação dos parâmetros/variáveis que se deseja, e calcular a correlação entre os fatores correspondentes extraídos das duas tabelas.

Exemplos:

- i) Em tabelas cujas entradas são notas inteiras aferidas por avaliadores pode-se querer testar o grau de subjetividade nos resultados modificando-se ligeiramente as entradas, como por exemplo: à nota n_{ij} associa-se outra retirada com lei uniforme do conjunto $\{n_{ij}-1, n_{ij}, n_{ij}+1\}$
- ii) Em tabelas cujas entradas são percentuais ou índices arredondados pode-se testar a interferência do arredondamento procedendo como no exemplo anterior. à n_{ij} associa-se o valor tirado com lei uniforme do intervalo contínuo $(n_{ij} - 0,5, n_{ij} + 0,5)$
- iii) Muitas vezes existe uma ou mais linhas (resp. colunas) que contribuem em demasia para um dado autovalor (relembremos dos capítulos anteriores que $\lambda_\alpha = \sum_i F_\alpha^2(i) f_i = \sum_j G_\alpha^2(j) f_j$),
perfaz-se então nova análise numa tabela sem esses elementos, que podem estar encobrindo relações entre os restantes. Em casos i e

ii tem-se verificado que o primeiro fator permanece praticamente inalterado, i.e. correlação próxima de 1 entre os dois respectivos, as mudanças, se existem, ocorrendo a partir do segundo. Isto é um fato mais ou menos esperado, uma vez que sabemos ser o primeiro eixo quem reflete a estrutura mais global, melhor dizendo robusta, da configuração. No terceiro exemplo, muitas vezes a não retirada mas o tratamento de tais elementos como suplementares é suficiente para realçar novas inter-relações.

4.2 2 Estudos Monte Carlo

Se se deseja obter a distribuição de algum parâmetro extraído de uma correspondência, a maneira mais simples é através métodos de simulação. Mantendo-se o tamanho da amostra, ou soma de todas as celas, pode-se gerar várias tabelas, a partir de uma dada inicialmente, usando-se a lei empírica f_{IJ} fornecida pela primeira. Com isso, pode-se obter informações distribucionais a respeito dos autovalores, dos fatores de cada nuvem, da contribuição de cada ponto, etc., ponto de partida para a obtenção de testes estatísticos dos resultados. Resultados extremos para um dado cardinal $|I|$ de linhas e $|J|$ de colunas são obtidos gerando-se cruzamentos nessas condições segundo a lei

$$\forall (i, j) \in I \times J, f_{ij} = \frac{1}{|I| \times |J|}$$

uma vez que ela nos dá a maior dispersão possível para uma tabela com essas restrições de cardinalidade nas margens. Sendo o qui-quadrado de tais tabelas bem baixo, os resultados médios obtidos com um dado tamanho de amostra servem de comparação extrema com os advindos de um particular estudo sob as mesmas restrições, a saber, tamanho de amostra e cardinalidade das categorias das margens.

Exemplo:

Em um estudo de simulação levado a cabo no laboratório de estatística de Paris ([2]) para uma tabela a 8 linhas e 8 colunas segundo a idéia acima, obteve-se uma quase invariância, com respeito a tamanho amostral, da taxa de variação explicada pelos cinco primeiros autovalores, bem como uma insensibilidade do erro relativo para os 3 primeiros autovalores (próximo de 1/4 para vários tamanhos de amostra).

4.2.3 Considerações estatísticas elementares

Um dos poucos testes estatísticos fáceis de se obter e de grande utilidade prática em correspondências é o relativo a hipóteses de significância ao nível α para as projeções nos diversos subespaços fato-

riais dos pontos de uma configuração $\{f_J^i - f_J\}_{i \in I}$ afetados das massas f_i . Notando que sob $H_0 : P_J^i = P_J \quad \forall i \in I$, a quantidade $n_i \|f_J^i - f_J\|^2$ é assintoticamente um $\chi^2_{|J|-1}$, temos, lembrando o Capítulo 2.

$$n_i \|f_J^i - f_J\|^2 = n \cdot x_{f_i} \cdot \sum_{\alpha \in A} F_\alpha^2(v)$$

Ora, o membro direito da fórmula acima nada mais é do que a contribuição relativa do j -ésimo ponto à inércia total, multiplicada pelo tamanho da amostra. Logo se $\chi^2_{|J|-1, \alpha}$ representa a abscissa de um qui-quadrado $|J|-1$ g.l. que deixa uma cauda de peso α temos que se o j -ésimo ponto da configuração for significativo ele deve estar no complementar da esfera de raio

$$\left(\frac{\chi^2_{|J|-1, \alpha}}{n_i} \right)^{1/2}$$

centrada no ponto f_J do R^J

A partir daí pode-se obter raios dos círculos de guarda da projeção do j -ésimo ponto nos diversos planos fatoriais; para o plano 1×2 por exemplo o raio vale:

$$\left(\frac{\chi^2_{2, \alpha}}{n_i} \right)^{1/2}$$

Outra consideração estatística útil é referente ao número de autovalores que devem ser extraídos para se obter uma explicação razoável da configuração. Esse número não deve ultrapassar um k dado pelo seguinte raciocínio:

sejam $\lambda_{\alpha 1} \geq \lambda_{\alpha 2} \geq \dots \geq \lambda_{\alpha k}$ os autovalores então k é tal que,

$$n \cdot \sum_{i,j} \frac{(f_{ij} - f_i \cdot f_j)^2}{f_i \cdot f_j} - \sum_{e=1}^{k-1} \lambda_{\alpha e} \geq (|I|-1)(|J|-1)$$

$$n \cdot \sum_{i,j} \frac{(f_{ij} - f_i \cdot f_j)^2}{f_i \cdot f_j} - \sum_{e=1}^k \lambda_{\alpha e} \leq (|I|-1)(|J|-1)$$

ou seja, após a extração do k -ésimo autovalor a inércia residual é menor do que a esperança matemática da inércia da nuvem que é o número de seus graus de liberdade, pois ela é um qui-quadrado.

Outros testes ou resultados dessa natureza dependem, como se sabe, da distribuição inicial a se considerar. Apresenta-se a área bem aberta à estudos com famílias de distribuições iniciais (tanto f_{IJ} como p_{IJ}) que permitam obter resultados realmente interessantes e dentro do espírito do método.

Capítulo 5

UMA VISÃO DAS APLICAÇÕES DA ANÁLISE DE CORRESPONDÊNCIAS

Pode-se dizer que a análise de correspondências está à espera da imaginação dos pesquisadores para encontrar novas aplicações para suas técnicas. Dentre os variados usos e inter-relacionamentos que ela tem gerado em estudos exploratórios nas áreas da sociologia, lingüística, economia, ecologia, psicologia, teoria da comunicação, medicina, geografia humana e econômica e outras mais, podemos destacar a título de exemplo, os tópicos seguintes:

5.1 Correspondências e teoria da informação

A partir da relação formal entre a entropia de uma lei de probabilidade P_{IJ} , sobre um espaço produto $I \times J$, com marginais P_I e P_J , e a distância qui-quadrado dessa mesma lei à lei produto correspondente, a saber:

$$H(P_{IJ}; P_I, P_J) = \sum_{i,j} P_i \cdot P_j \varphi \left(\frac{P_{ij}}{P_i \cdot P_j} \right)$$

$$d(P_{IJ}, P_{I \times J}) = \sum_{i,j} P_i \cdot P_j \Psi \left(\frac{P_{ij}}{P_i \cdot P_j} \right)$$

onde $\varphi, \Psi: \mathbf{R}_+ \rightarrow \mathbf{R}$ sendo $\varphi(x) = x \log_2 x$

$$\text{e } \Psi(x) = x^2 - 1$$

importantes analogias podem ser extraídas. Em particular, certos algoritmos e conceitos da teoria da informação possuem uma contrapartida em correspondências que muitas vezes substitui convenientemente a utilização dos primeiros. A taxionomia e o reconhecimento de formas, que muito utilizam os conceitos da teoria da informação, têm aplicado com relativo sucesso a análise de correspondências para obter os mesmos resultados que eles possibilitam.

5.2 Correspondências e classificação hierárquica

O problema de se obter classificações hierárquicas dado um cruzamento de propriedades por elementos de um conjunto desafia até hoje estatísticos e taxionomistas. Oriundos de várias idéias e disciplinas tem-se atualmente um número considerável de algoritmos distintos, todos

eles impotentes quando o cardinal do conjunto a ser classificado ultrapassa a ordem de algumas centenas. A partir da analogia entre a fórmula de reconstituição de uma tabela cruzada, via correspondência, e a expansão ortogonal de uma medida de probabilidade sobre o produto cartesiano de um conjunto, munido de uma hierarquia binária indiciada, por ele mesmo, pode-se utilizar a análise de correspondências para detectar os ramos iniciais de uma hierarquia binária em conjuntos de cardinal elevado. A importância disso para estudos de classificação em diversas áreas das ciências humanas e sociais é evidente ([2], [], []).

5.3 Correspondências e consistência estatística de dados

Tais estudos tem se desenvolvido ultimamente em virtude do aumento de complexidade do problema da crítica de dados para o “data base” de grandes centros de informática. Técnicas multivariadas tem sido empregadas obtendo-se em geral testes de hipóteses baseados nos valores de certas estatísticas apropriadas (ver p. ex., [11]). Ora pela força de sua saída gráfica, a análise de correspondências desempenha para o problema de examinar a consistência de um levantamento em grande escala de várias variáveis inter-relacionadas papel semelhante ao do histograma quando se examina uma única amostra numérica. Em grandes esforços censitários, como o ENDEF ora realizado no País, o método tem sido testado para confirmar o surgimento de certas estruturas básicas que devem ser retratadas pelos questionários.

5.4 “Scaling”, “parametric representation” e correspondências

Conforme dito no Capítulo 0, o método aqui apresentado tem conseguido unir ou substituir várias técnicas relativamente isoladas de análise multivariada. Em particular ela propõe uma alternativa à famosa técnica de *scaling*, introduzida por Shepard e idéias profícuas no campo da “parametric representation” de Kruskal & Carroll ([2] [14]).

5.5 Grafos e correspondências

É fato que em várias disciplinas sociais e em problemas relativos ao armazenamento de informações a teoria dos grafos tem contribuído de forma considerável com linguagem, metodologia e resultados. A partir da análise de correspondências da matriz de incidência de um grafo conclusões e propostas interessantes foram levantadas pelo grupo de Paris VI ([2]). Em particular, considerando-se as tabelas simétricas como matrizes de incidência de grafos “ponderados” certas interpretações obtidas da correspondência efetuada em grafos típicos foram extendidas para tabelas simétricas em geral por Ludovic Lebart.

5.6 A avaliação estatística de sistemas de computação e a análise de correspondências

Os trabalhos de Sharpe [21] e Drummond [4] servem para dar uma idéia do estado de coisas nesse recente campo, onde vários modelos distintos subexistem. Para a obtenção de figuras de mérito e fórmulas que avaliem a performance dos diversos elementos de um moderno sistema de computação acreditamos que a análise de correspondências venha a ser de grande utilidade.

PARTE II

INTRODUÇÃO

Apresentamos aqui alguns exemplos concretos de saída em análise de correspondências procurando ressaltar tipos diversos de aplicações. Para se iniciar na prática da interpretação dos resultados, aconselhamos que em cada exemplo, após se ler e compreender a estrutura e o porque do cruzamento montado, a análise seja começada pela saída gráfica, devido a seu forte apelo analógico. Apesar de haver nos exemplos uma insinuação didática de como se deve abordar o resultado (se bem que não haja regra geral para isso) gostaríamos de ressaltar aqui alguns poucos parâmetros numéricos, cuja discussão foi feita no Capítulo 2 da Parte I, que podem servir de guia.

Contribuição relativa do eixo à inércia

É o valor da razão entre o autovalor respectivo e a inércia total da nuvem, que é o traço da matriz a ser diagonalizada. Este parâmetro é interessante para se ter uma idéia da importância do eixo como representação da nuvem. Geralmente a partir de no máximo o quinto eixo a contribuição passa a ser insignificante.

Contribuição relativa do ponto à inércia total

É o valor da inércia do ponto em relação ao centro de gravidade da nuvem dividido pela inércia total da nuvem. Um ponto com uma contribuição baixa ou está muito próximo do centro de gravidade ou possui massa pequena, isso significa que a linha/coluna que ele representa é

ou um tipo não diferenciado pela tabela ou cuja ocorrência é rara na população. Geralmente deve-se refazer a análise colocando-se esses elementos como suplementares ou simplesmente retirando a linha/coluna correspondente da tabela. Na saída do programa este valor aparece para cada ponto, multiplicado por 1.000, sob o nome de INÉRCIA.

Contribuição relativa do ponto à inércia do eixo

Analogamente ao caso anterior, modificando-se o denominador para a inércia do eixo. Mede pois a importância que deve ser dada ao ponto quando examinamos o eixo em questão. Na saída do programa esses valores aparecem multiplicados por 1.000 na coluna de nome CTR RELATIVA.

Correlação do ponto centrado (i.e. referido ao CG da nuvem) com o eixo

Como se sabe da estatística, é uma medida da “proximidade geométrica” do ponto/vetor ao eixo. O conjunto de pontos com os quais o eixo possui uma correlação alta, pode nos dar como em análise fatorial uma interpretação para o eixo. Note-se porém que um ponto pode apresentar uma correlação alta porém uma baixa contribuição à inércia do eixo (veja Figura 1).

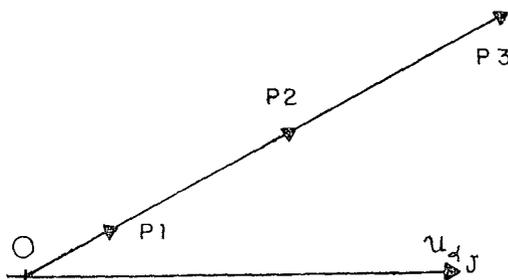


FIGURA 1

Os pontos P1, P2 e P3 têm mesma massa. Todos possuem a mesma correlação com u_{1j} , porém a contribuição à inércia decresce de P3 para P1.

Aparecem multiplicados por 1.000 na coluna COR da saída do programa.

Correlação do ponto com o plano fatorial

Mede agora a proximidade geométrica do ponto ao plano fatorial. Vale a mesma ressalva feita no item anterior.

Generalizando o que foi dito nos dois parágrafos anteriores é interessante observar para cada ponto a proximidade com sua projeção no maior subespaço utilizado para as interpretações. Devido à ortogonalidade dos eixos a soma dos quadrados dos coeficientes de correlação com os eixos que geram o subespaço é uma medida desse fato. Tal valor aparece na saída ($\times 1.000$) sob o nome de QLT.

Nota sobre o programa

Para uso livre e eficiente do método de correspondências é necessário possuí-lo programado em computador, uma vez que a sua execução manual é por demais onerosa, senão impraticável, quando o tamanho da tabela cresce. O programa utilizado para a elaboração desses exemplos encontra-se implantado no Sistema de Análise de Dados e Estatística — SADE, do CI. É capaz de aplicar o método a cruzamentos cujo cardinal das linhas com as colunas chegue até 400. Fornece as projeções de uma ou ambas as nuvens em diversos planos fatoriais e os principais parâmetros numéricos discutidos anteriormente.

Nota sobre a saída gráfica

Quem leu atentamente a Parte I e por acaso folheie uma das saídas gráficas aqui apresentadas poderá ter se surpreendido com o fato de que os diagramas estão representados em um plano com um sistema ortogonal de coordenadas, enquanto que o sistema de coordenadas inicial, ortogonal segundo qui-quadrado, não o é. A explicação é simples: suponha-se que no plano $u_{1J} \times u_{2J}$ temos o ponto i representado pelo par $(F_1(i), F_2(i))$ e:

$$\text{proj}_{u_{1J} \times u_{2J}} i = F_1(i) u_{1J} + F_2(i) u_{2J}$$

Conforme se ilustrou no parágrafo 3.1, Parte I, a transformação linear:

$S: R^J \rightarrow R^J$, dada pela matriz $(S)_{jj'} = S_{jj'} f_j^{-1/2}$ relativa à base canônica, leva vetores ortogonais qui-quadrado em vetores ortogonais euclides:

$$\begin{aligned} \langle Su_{1J}, Su_{2J} \rangle &= \sum_j \frac{u_{1j} u_{2j}}{f_j} = \langle u_{1J}, u_{2J} \rangle \chi^2 = 0 \\ \langle Su_{1J}, Su_{1J} \rangle &= \sum_j \frac{u_{1j}^2}{f_j} = 1 \end{aligned}$$

Logo o ponto i ficará:

$$S(\text{proj}_{u_{1J} \times u_{2J}} i) = F_1(i) Su_{1J} + F_2(i) Su_{2J}$$

assim o plano que aparece nos gráficos é o referido aos eixos ortonormais $\{Su_{1J}, Su_{2J}\}$ e não existe alteração nenhuma da estrutura da nuvem uma vez que os valores de suas coordenadas (= fatores) continuam inalterados, e são eles os elementos que entram em jogo nas fórmulas e princípios do método

Capítulo 1

A POSIÇÃO SOCIAL DAS OCUPAÇÕES *

1.1 Neste exemplo utilizamos o trabalho de Nelson do Valle [24], que mediante uma série de análises criou um conjunto de índices de “status”

* O arquivo para essas análises foi mostrado por Cláudio Mendes de Oliveira

social para as ocupações discriminadas no Censo Demográfico de 70. Em particular, do Valle utilizou como determinantes principais de "status" a renda do indivíduo e seu nível de escolaridade. Empregou ele uma amostra de 1/28 da população economicamente ativa do Estado de Guanabara na faixa de idade entre 15 e 64 anos, extremos inclusive. Isso acarretou que duas ocupações da classificação censitária não aparecessem na amostra.

O índice de do Valle é crescente com o "status", isto é, com a posição social expressa basicamente por uma combinação de renda e instrução. Maiores detalhes podem ser encontrados na publicação citada bem como no exemplo do manual SADE, versão 1 [20]. Para nós interessa o fato que tal índice é uma escala sociológica. Assim sendo, agrupamos os índices em 19 classes em ordem crescente com a dos índices. Isso foi feito primeiro identificando as ocupações que caíssem em faixas de comprimento 1, iniciando com [1,50; 2,50) e terminando com o intervalo [88,50; 89,50) pois o mais baixo índice é 1,81 e o mais alto 88,75. Após agregamos os grupos assim obtidos em 19 outros maiores, levando agora em conta não só a contigüidade como também o número de pessoas em cada grupo/intervalo. Apesar disso, os dois primeiros grupos apresentam-se com bem mais indivíduos que os restantes.

1.2 A partir daí efetuamos um cruzamento dos 19 grupos com três variáveis, renda, educação e sexo, assim categorizadas:

VARIÁVEL	CÓDIGO	
renda	AI	rendimentos acima de 0 até 100 cruzeiros, inclusive
	BI	rendimentos de 101 a 150 cruzeiros, inclusive;
	CI	rendimentos de 151 a 200 cruzeiros, inclusive;
	DI	rendimentos de 201 a 500 cruzeiros, inclusive;
	EI	rendimentos iguais a ou acima de 501 cruzeiros;
educação	FI	nível educacional igual ao ciclo elementar complementar;
	HI	nível educacional igual ao 1.º ciclo médio completo
	II	nível educacional igual ao 2.º ciclo médio ou curso superior, completos;
sexo	JI	masculino;
	KI	feminino.

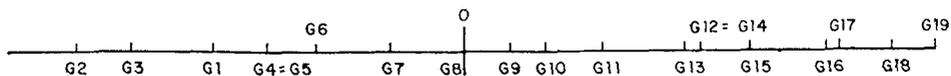
Tomamos a população economicamente ativa da Guanabara, na faixa etária idêntica a de do Valle, e efetuamos o cruzamento dessas variáveis com os grupos ocupacionais agregados, conforme descrito agora: Cada pessoa contribuiu na linha respectiva a seu grupo ocupacional com uma unidade em cada cela cuja categoria correspondentemente ela possuía. Ou, de outra forma, cada cela do cruzamento nos dá

número de pessoas no determinado grupo ocupacional que está na particular categoria. Ora, uma vez que as variáveis escolhidas foram os determinantes da classificação das linhas (a menos do sexo, mas que podemos considerar, ainda hoje, como um determinante ocupacional), é de se esperar que ao se submeter o cruzamento a uma correspondência *apareça efeito Guttman no conjunto das linhas*. Se a classificação de do Valle é boa, e se o agrupamento segundo os índices foi bom, os dezenove pontos devem aparecer segundo uma parábola no plano dos dois primeiros eixos fatoriais, e sua projeção no primeiro eixo nos deve dar uma escala cuja ordem *coincida com a ordem dos grupos*.

Submeteu-se o cruzamento à rotina de correspondências desenvolvida no C.I. e obtivemos os resultados apresentados nas Figuras 1 e 2.

1.3 Plano dos dois primeiros eixos fatoriais

Encontram-se aí apresentados os dois conjuntos, a saber, linhas e colunas. Os grupos ocupacionais apresentam nitidamente efeito Guttman, com a execução de G1, que se apresenta bastante deslocado, e um pequeno desvio de G12. Projetando-se os pontos no primeiro eixo, obtemos a escala a seguir apresentada:

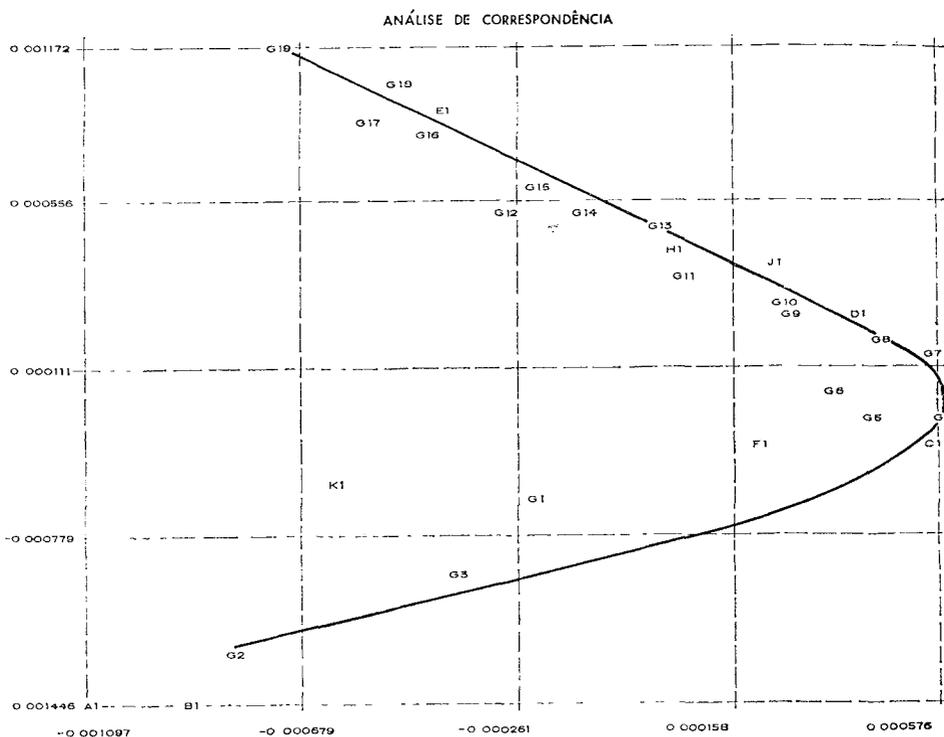


O deslocamento do grupo 1, deve-se provavelmente a uma falha no agrupamento, pois como já se disse os dois primeiros grupos e em particular o primeiro apresentam cardinal mais elevado que o restante. Note-se que as variáveis classificatórias apresentam-se bem entremeadas ao longo da parábola. Assim, as duas classes de renda mais baixa encontram-se no canto inferior esquerdo, próximas ao início da parábola embora no primeiro eixo suas projeções coincidam. As duas outras, C1 e D1, estão no aglomerado das ocupações de “status” médio, no bojo da parábola. A classe D1 em especial, é aproximadamente a renda média dos 19 grupos ocupacionais. Realmente, o próximo grupo já mais afastado de D1 ao longo da parábola, o G11, contém entre outras ocupações os aviadores civis, contadores, veterinários e agrônomos o que implica num certo salto social com relação aos 10 anteriores que agrupavam desde as ocupações elementares até as de médio escalão burocrático como taquígrafos, escrivães, agentes fiscais, ou de caráter técnico muito específico, como meteorologista [7], [24]. Os grupos de maior “status” acham-se envolvidos em um compromisso entre a maior classe de renda e a mais elevada categoria educacional (II). As ocupações de G19 e G18, que envolvem a classe dos magistrados, procuradores, promotores, médicos e engenheiros acham-se mais próximas de II que E1 no primeiro eixo. Na realidade o “status” que possuem é mais devido ao respeito de ordem moral ou técnica que infundem, ou o poder que possuem, do que aos rendimentos que percebem. Em grupos inferiores existem muitas ocupações com rendimentos superiores às antes descritas [7], [24] que por não estarem envolvidas pela mesma “aura de saber” não

têm o mesmo nível social. Os homens (I1) acham-se envolvidos no ramo ascendente da parábola enquanto as mulheres (K1), embora não caracterizando grupo algum apresentam-se mais próximas à parte descendente da parábola. As duas categorias do sexo se encontram bem separadas, as mulheres, confirmando sua rarefação e quase ausência nos grupos altos acham-se nos dois eixos atraídas por G2 e G3, ao passo que os homens encontram-se numa posição bem central relativamente à nuvem ocupacional.

FIGURA 1

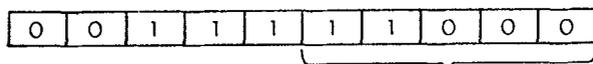
A projeção de ambas as nuvens, no primeiro plano fatorial 1×2



1.4 Com o objetivo de testar novamente o efeito Guttman cruzamos os mesmos grupos ocupacionais anteriores com as 10 colunas de um questionário desdobrado de Guttman assim criado:

- 1.^a coluna/pergunta Sua renda é maior ou igual a 1501?
- 2.^a coluna/pergunta: Sua renda é maior ou igual a 501?
- 3.^a coluna/pergunta: Sua renda é maior ou igual a 201?
- 4.^a coluna/pergunta: Sua renda é maior ou igual a 101?
- 5.^a coluna/pergunta: Sua renda é maior ou igual a 51?

As cinco colunas restantes sendo o complementar das anteriores, i.e., para cada indivíduo da população economicamente ativa da GB entre 15 e 64 anos coloca-se na linha correspondente à sua ocupação uma unidade em cada uma das cinco primeiras celas em que sua resposta for sim e exatamente o inverso nas cinco seguintes. Obviamente essa tabela aproxima uma de Guttman, pois, por exemplo, uma pessoa com renda igual a 490 contribuiria assim para a sua linha:



complementar da seqüência anterior,

além do mais o início da seqüência de uns obviamente desloca-se para a direita à medida que decresce a ordem dos grupos ocupacionais.

1.5 Nas Figuras 3, 4 e 5 temos a projeção da nuvem ocupacional nos planos fatoriais 1×2 , 1×3 e 1×4 . Repare-se no primeiro plano novamente a existência do efeito Guttman (a menos dos grupos 2 e 12) sendo a assimetria da curva devido à maior concentração de pessoas na rama esquerda da parábola. Nos planos seguintes deve aparecer a sucessão de polinômios ortogonais (vide Cap. 4, Parte I). Em 1×3 temos realmente o segundo polinômio ortogonal com suas 3 raízes e 2 pontos extremos. No plano seguinte o desenho do polinômio já é mais obscuro devido ao fato da variância no quarto eixo ser extremamente pequena causando quase que uma ordenação visual linear dos dados no plano 1×4 . Recorremos então ao ajuste de um "spline" cúbico pelos pontos obtendo-se o resultado bem satisfatório documentado na Figura 6. Obtém-se ainda nesse plano, juntamente com a observação de 2 anteriores, uma perfeita noção das intensidades das diferenças na escala obtida:

- os três primeiros grupos, extremamente próximos na escala (=1.º eixo), apresentam fortes diferenças entre si; provavelmente isto é devido ao fato deles concentrarem as ocupações de nível extremamente baixo (aprendizes — 3,31, engraxates — 3,76, lixeiros — 5,56, tropeiros — 4,61, bordadeiras e cerzideiras — 4,41, chacareiros, hortelões e floricultores — 5,04, etc) mas que apresentam uma certa diversidade de renda entre inclusive elementos da mesma ocupação/grupo, discriminadas pelas perguntas da 3.ª, 4ª e 5ª colunas. Além do mais, como se disse, concentraram eles grande parte da população;
- segue-se após um conjunto homogêneo de G4 a G9 para depois os saltos serem praticamente constantes de G11/G10 a G12, G12 a G14/G15 e G13, deles a G16, deste a G17 e de G17 a G18, destacando-se nitidamente o grupo 19 conforme razões ventiladas anteriormente.

Comparando-se esse escalograma com o obtido na Figura 1, onde outras variáveis classificatórias foram usadas, parece-nos que os seguintes aglomerados ("clusters") ocupacionais ficam bem definidos:

um composto de G1, G2 e G3;

um composto de G4, G5, G6 e G7;

G8;

G9;

G10 e G11;

G12;

G13;

G14 e G15;

e os restantes sem alteração.

Caso fosse de interesse do pesquisador determinar menos aglomerados ocupacionais sugerir-se-ia então uma nova rodada do método com esses grupos redefinidos. A análise de dados é uma técnica interativa cujo ponto de parada é responsabilidade única e exclusiva do pesquisador. É nada tão dentro de seu espírito como a análise de correspondências que permite melhorar a cada passo resultados obtidos ou modificar a definição do próprio arquivo de dados para que se torne mais adequado aos objetivos desejados.

Os resultados numéricos relativos à projeção no plano 1x2

***** PAGINA 1

***** ANALISE DE CORRESPONDENCIA *****

RESULTADOS FATORIAIS PARA O CONJUNTO DAS LINHAS						
NOME DA VARIÁVEL	CATEGORIA		FATOR			
	IN	DESCRICAÇÃO	PROJECAO	CORRELACAO	CTR. RELATIVA	
GRUPO	1	G1	-0.000689	0.000615	0.000000	
	2	G2	-0.001299	0.000808	0.000543	
	3	G3	-0.000976	0.000902	0.000026	
	4	G4	-0.000386	0.000212	0.000014	
	5	G5	-0.000347	0.000376	0.000013	
	6	G6	-0.000286	0.000394	0.000050	
	7	G7	-0.000111	0.000037	0.000001	
	8	G8	-0.000061	0.000017	0.000001	
	9	G9	0.000038	0.000011	0.000000	
	10	G10	0.000104	0.000112	0.000005	
	11	G11	0.000192	0.000216	0.000004	
	12	G12	0.000429	0.000685	0.000054	
	13	G13	0.000381	0.000793	0.000025	
	14	G14	0.000473	0.000793	0.000004	
	15	G15	0.000568	0.000885	0.000036	
	16	G16	0.000750	0.000839	0.000122	
	17	G17	0.000799	0.000743	0.000046	
	18	G18	0.000953	0.000836	0.000058	
	19	G19	0.001172	0.000711	0.000119	

***** PAGINA 2

***** ANALISE DE CORRESPONDENCIA *****

RESULTADOS FATORIAIS PARA O CONJUNTO DAS COLUNAS						
NOME DA VARIÁVEL	CATEGORIA		FATOR			
	INUMERO	DESCRICAÇÃO	PROJECAO	CORRELACAO	CTR. RELATIVA	
F01	1	A1	-0.001498	0.000718	0.000169	
F02	1	B1	-0.001496	0.000806	0.000131	
F03	1	C1	-0.000479	0.000358	0.000046	
F04	1	D1	0.000066	0.000020	0.000002	
F05	1	E1	0.000847	0.000890	0.000230	
F06	1	F1	-0.000464	0.000876	0.000130	
F07	1	H1	0.000283	0.000355	0.000016	
F08	1	I1	0.001024	0.000754	0.000229	
F09	1	J1	0.000258	0.000563	0.000048	
F10	1	K1	-0.000640	0.000563	0.000120	

RESULTADOS FATORIAIS PARA O CONJUNTO DAS LINHAS							
NOME DA VARIAVEL	CATEGORIA		FATOR				
	INUMERO	DESCRICA	PROJECAO	CORRELACAO	CTR. RELATIVA		
GRUPO	1	G1	-0.000212	0.000058	0.000000		
	2	G2	-0.000806	0.000311	0.000395		
	3	G3	-0.000366	0.000127	0.000007		
	4	G4	0.000576	0.000470	0.000059		
	5	G5	0.000448	0.000627	0.000041		
	6	G6	0.000369	0.000657	0.000158		
	7	G7	0.000566	0.000948	0.000055		
	8	G8	0.000456	0.000970	0.000059		
	9	G9	0.000283	0.000619	0.000009		
	10	G10	0.000263	0.000718	0.000061		
	11	G11	0.000062	0.000022	0.000001		
	12	G12	-0.000272	0.000277	0.000041		
	13	G13	0.000027	0.000004	0.000000		
	14	G14	-0.000132	0.000062	0.000001		
	15	G15	-0.000217	0.000130	0.000010		
	16	G16	-0.000427	0.000272	0.000075		
	17	G17	-0.000543	0.000343	0.000040		
	18	G18	-0.000481	0.000212	0.000028		
	19	G19	-0.000706	0.000257	0.000082		

RESULTADOS FATORIAIS PARA O CONJUNTO DAS COLUNAS							
NOME DA VARIAVEL	CATEGORIA		FATOR				
	INUMERO	DESCRICA	PROJECAO	CORRELACAO	CTR. RELATIVA		
F01	1	A1	-0.001097	0.000385	0.000171		
F02	1	B1	-0.000888	0.000284	0.000087		
F03	1	C1	0.000559	0.000489	0.000118		
F04	1	D1	0.000414	0.000769	0.000119		
F05	1	E1	-0.000402	0.000200	0.000098		
F06	1	F1	0.000224	0.000204	0.000057		
F07	1	H1	0.000055	0.000013	0.000001		
F08	1	I1	-0.000666	0.000319	0.000184		
F09	1	J1	0.000245	0.000505	0.000082		
F10	1	K1	-0.000606	0.000505	0.000204		

FIGURA 3

Saída numérica da projeção da nuvem dos grupos no plano
1x2

***** PAGINA 1

***** ANALISE DE CORRESPONDENCIA *****

RESULTADOS FATORIAIS PARA O CONJUNTO DAS LINHAS							
NOME DA VARIÁVEL	CATEGORIA		FATOR				
	INUMERO	DESCRICA	PROJEC	CORRELAC	CTR. RELATIVA		
GRUPO	1	G1	-0.000351	0.000387	0.000000		
	2	G2	-0.000415	0.000309	0.000108		
	3	G3	-0.000398	0.000278	0.000008		
	4	G4	-0.000242	0.000542	0.000011		
	5	G5	-0.000212	0.000443	0.000010		
	6	G6	-0.000195	0.000495	0.000045		
	7	G7	-0.000165	0.000388	0.000005		
	8	G8	-0.000131	0.000365	0.000005		
	9	G9	-0.000093	0.000204	0.000001		
	10	G10	-0.000051	0.000105	0.000002		
	11	G11	-0.000056	0.000100	0.000001		
	12	G12	0.000088	0.000311	0.000004		
	13	G13	0.000238	0.000569	0.000019		
	14	G14	0.000188	0.000408	0.000001		
	15	G15	0.000188	0.000346	0.000008		
	16	G16	0.000424	0.000633	0.000076		
	17	G17	0.000610	0.000659	0.000052		
	18	G18	0.000791	0.000648	0.000078		
	19	G19	0.001184	0.000583	0.000237		

***** PAGINA 2

***** ANALISE DE CORRESPONDENCIA *****

RESULTADOS FATORIAIS PARA O CONJUNTO DAS COLUNAS							
NOME DA VARIÁVEL	CATEGORIA		FATOR				
	INUMERO	DESCRICA	PROJEC	CORRELAC	CTR. RELATIVA		
F01	1	A1	0.001558	0.000565	0.000135		
F02	1	B1	0.001388	0.000604	0.000184		
F03	1	C1	0.000701	0.000610	0.000184		
F04	1	D1	0.000059	0.000177	0.000004		
F05	1	E1	0.000009	0.000163	0.000000		
F06	1	F1	-0.000076	0.000565	0.000007		
F07	1	H1	-0.000121	0.000604	0.000016		
F08	1	I1	-0.000321	0.000610	0.000084		
F09	1	J1	-0.000743	0.000177	0.000049		
F10	1	K1	-0.000806	0.000163	0.000008		

***** ANALISE DE CORRESPONDENCIA *****

RESULTADOS FATORIAIS PARA O CONJUNTO DAS LINHAS

NOME DA VARIÁVEL	CATEGORIA		FATOR		
	INUMERO	DESCRICA	PROJECAO	CORRELACAO	CTR. RELATIVA
GRUPO	1	G1	-0.000245	0.000189	0.000000
	2	G2	-0.000448	0.000360	0.000421
	3	G3	-0.000425	0.000316	0.000032
	4	G4	0.000051	0.000024	0.000002
	5	G5	0.000107	0.000113	0.000008
	6	G6	0.000087	0.000098	0.000030
	7	G7	0.000125	0.000222	0.000009
	8	G8	0.000117	0.000292	0.000013
	9	G9	0.000140	0.000465	0.000008
	10	G10	0.000119	0.000562	0.000043
	11	G11	0.000131	0.000555	0.000011
	12	G12	0.000012	0.000005	0.000000
	13	G13	0.000059	0.000035	0.000004
	14	G14	0.000088	0.000090	0.000001
	15	G15	0.000093	0.000085	0.000006
	16	G16	0.000004	0.000000	0.000000
	17	G17	-0.000064	0.000007	0.000002
	18	G18	-0.000150	0.000023	0.000009
	19	G19	-0.000359	0.000054	0.000073

***** ANALISE DE CORRESPONDENCIA *****

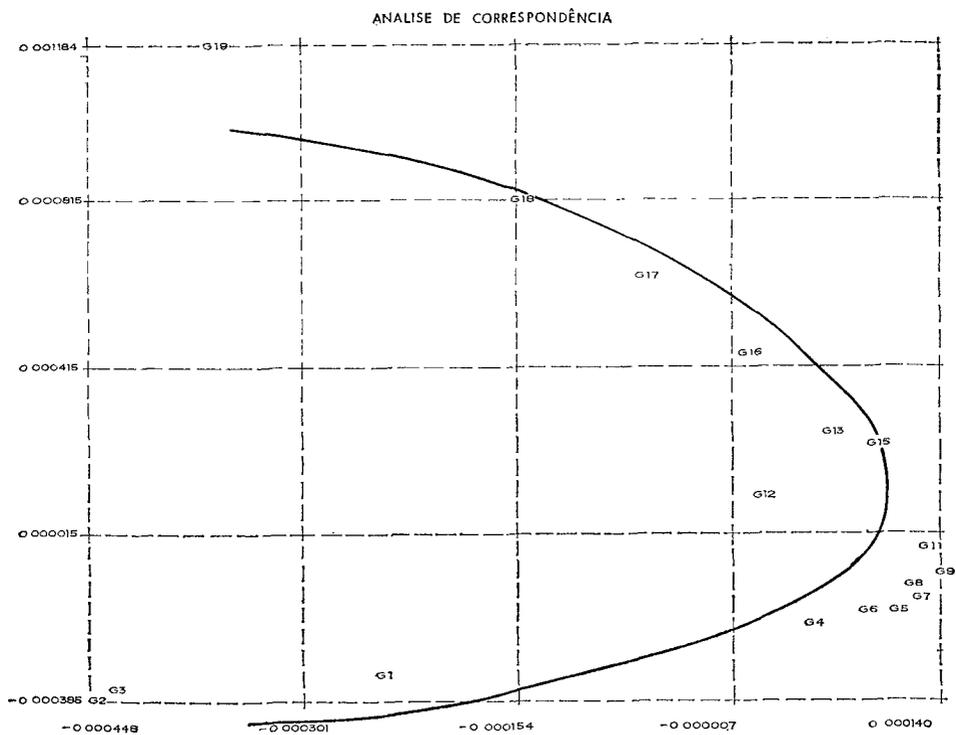
RESULTADOS FATORIAIS PARA O CONJUNTO DAS COLUNAS

NOME DA VARIÁVEL	CATEGORIA		FATOR		
	INUMERO	DESCRICA	PROJECAO	CORRELACAO	CTR. RELATIVA
F01	1	A1	-0.000528	0.000065	0.000052
F02	1	B1	-0.000408	0.000052	0.000053
F03	1	C1	0.000015	0.000000	0.000000
F04	1	D1	0.000098	0.000483	0.000035
F05	1	E1	0.000015	0.000459	0.000001
F06	1	F1	0.000026	0.000065	0.000003
F07	1	H1	0.000036	0.000052	0.000005
F08	1	I1	-0.000007	0.000000	0.000000
F09	1	J1	-0.001228	0.000483	0.000444
F10	1	K1	-0.001352	0.000459	0.000079

NCARAT= 3; 11 29
N1, N2

FIGURA 3 - A

Saída gráfica da projeção da nuvem dos grupos no plano 1x2



Saída numérica da projeção dos grupos no plano 1x3

***** PAGINA 9
 ***** ANALISE DE CORRESPONDENCIA *****

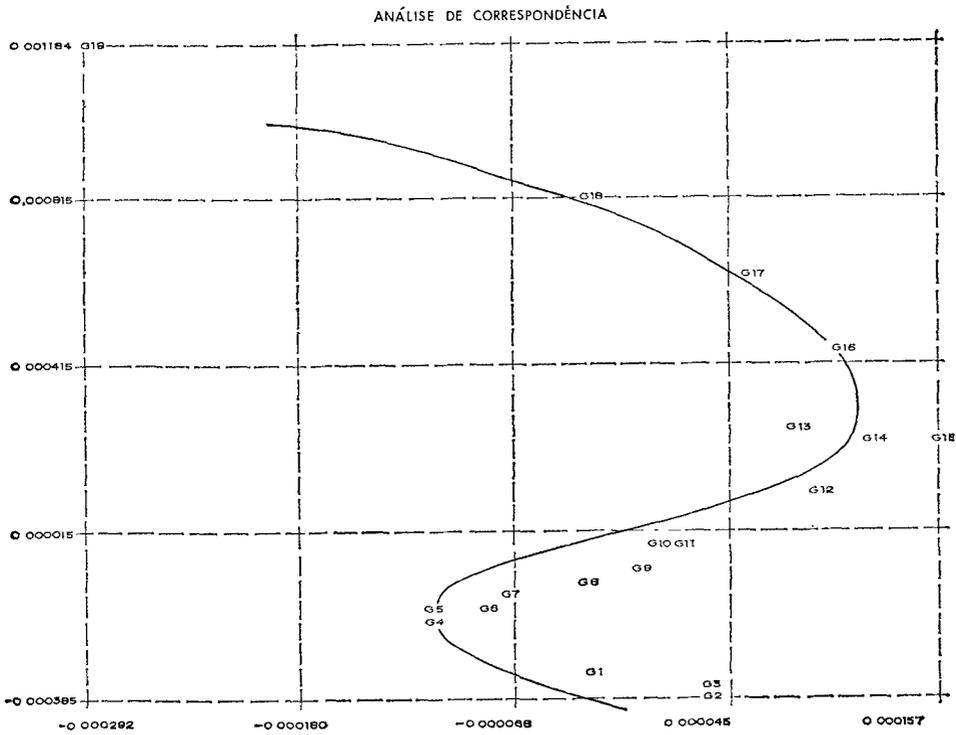
RESULTADOS FATORIAIS PARA O CONJUNTO DAS LINHAS						
NOME DA VARIÁVEL	CATEGORIA		PROJECAO	FATOR		
	INUMERO	DESCRICA		CORRELACAO	CTR. RELATIVA	
GRUPO	1	G1	-0.000023	0.000002	0.000000	
	2	G2	0.000037	0.000002	0.000013	
	3	G3	0.000038	0.000003	0.000001	
	4	G4	-0.000106	0.000105	0.000030	
	5	G5	-0.000108	0.000115	0.000036	
	6	G6	-0.000078	0.000079	0.000105	
	7	G7	-0.000066	0.000063	0.000011	
	8	G8	-0.000026	0.000015	0.000003	
	9	G9	0.000002	0.000000	0.000000	
	10	G10	0.000009	0.000003	0.000001	
	11	G11	0.000022	0.000015	0.000001	
	12	G12	0.000093	0.000348	0.000071	
	13	G13	0.000082	0.000068	0.000033	
	14	G14	0.000122	0.000172	0.000007	
	15	G15	0.000157	0.000240	0.000077	
	16	G16	0.000104	0.000038	0.000066	
	17	G17	0.000056	0.000006	0.000006	
	18	G18	-0.000026	0.000001	0.000001	
	19	G19	-0.000292	0.000035	0.000209	

***** PAGINA 10
 ***** ANALISE DE CORRESPONDENCIA *****

RESULTADOS FATORIAIS PARA O CONJUNTO DAS COLUNAS						
NOME DA VARIÁVEL	CATEGORIA		PROJECAO	FATOR		
	INUMERO	DESCRICA		CORRELACAO	CTR. RELATIVA	
F01	1	A1	-0.000424	0.000042	0.000146	
F02	1	B1	-0.000221	0.000015	0.000068	
F03	1	C1	0.000223	0.000062	0.000270	
F04	1	D1	-0.000015	0.000011	0.000003	
F05	1	E1	-0.000002	0.000007	0.000000	
F06	1	F1	0.000021	0.000042	0.000007	
F07	1	H1	0.000019	0.000015	0.000006	
F08	1	I1	-0.000102	0.000062	0.000124	
F09	1	J1	0.000183	0.000011	0.000043	
F10	1	K1	0.000170	0.000007	0.000005	

FIGURA 4 — A

Saída gráfica da projeção dos grupos no plano 1×3



Saída numérica no plano 1x4

***** PAGINA 15
 ***** ANALISE DE CORRESPONDENCIA *****

RESULTADOS FATORIAIS PARA O CONJUNTO DAS LINHAS							
NOME DA VARIÁVEL	CATEGORIA		FATOR				
	INUMERO	DESCRICAÇÃO	PROJECAO	CORRELACAO	CTR. RELATIVA		
GRUPO	1	G1	-0.000173	0.000094	0.000001		
	2	G2	-0.000015	0.000000	0.000038		
	3	G3	0.000207	0.000075	0.000592		
	4	G4	0.000011	0.000001	0.000006		
	5	G5	-0.000004	0.000000	0.000001		
	6	G6	-0.000001	0.000000	0.000000		
	7	G7	0.000001	0.000000	0.000000		
	8	G8	-0.000001	0.000000	0.000000		
	9	G9	0.000002	0.000000	0.000000		
	10	G10	0.000001	0.000000	0.000000		
	11	G11	0.000001	0.000000	0.000000		
	12	G12	-0.000013	0.000007	0.000026		
	13	G13	0.000004	0.000000	0.000001		
	14	G14	0.000004	0.000000	0.000000		
	15	G15	0.000006	0.000000	0.000002		
	16	G16	0.000005	0.000000	0.000003		
	17	G17	0.000003	0.000000	0.000000		
	18	G18	0.000001	0.000000	0.000000		
	19	G19	-0.000003	0.000000	0.000000		

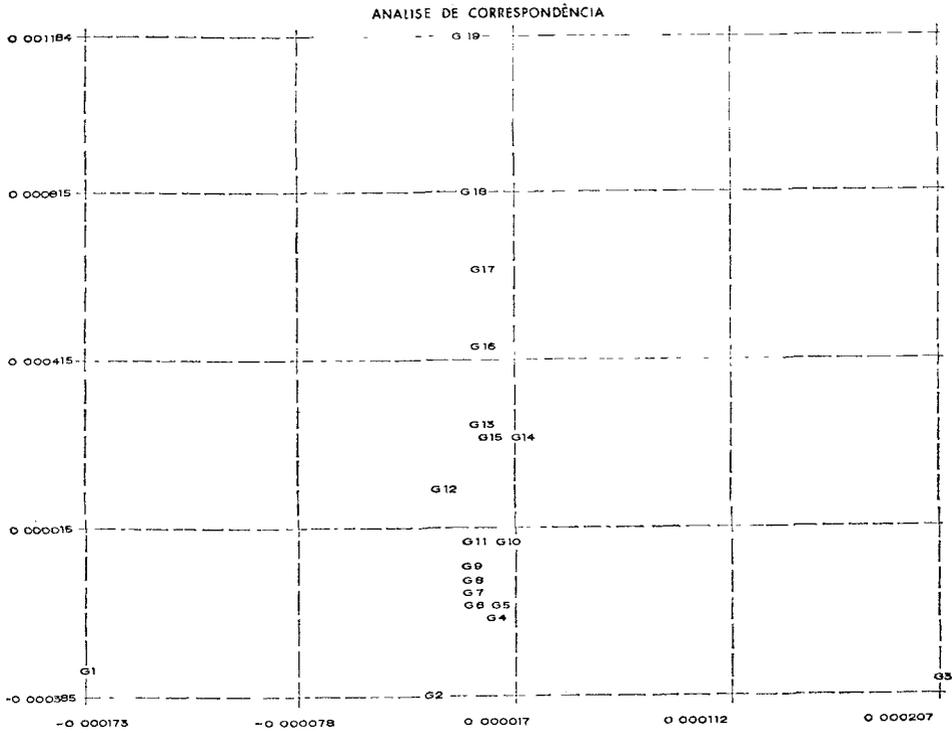
***** PAGINA 16
 ***** ANALISE DE CORRESPONDENCIA *****

RESULTADOS FATORIAIS PARA O CONJUNTO DAS COLUNAS							
NOME DA VARIÁVEL	CATEGORIA		FATOR				
	INUMERO	DESCRICAÇÃO	PROJECAO	CORRELACAO	CTR. RELATIVA		
F01	1	A1	-0.000004	0.000000	0.000000		
F02	1	B1	0.000001	0.000000	0.000000		
F03	1	C1	0.000001	0.000000	0.000000		
F04	1	D1	0.000005	0.000001	0.000007		
F05	1	E1	-0.000005	0.000042	0.000006		
F06	1	F1	0.000000	0.000000	0.000000		
F07	1	H1	-0.000000	0.000000	0.000000		
F08	1	I1	-0.000000	0.000000	0.000000		
F09	1	J1	-0.000062	0.000001	0.000087		
F10	1	K1	0.000410	0.000042	0.000571		

NCARAT= 3; 11 29
 N1.N2

FIGURA 5 — A

Saída gráfica no plano 1×4



Os resultados do ajuste de um spline cúbico

FUNÇÃO SPLINE	DERIVADA PRIMEIRA	DERIVADA SEGUNDA
0.038741	—1.574172	12.225573
—0.027047	—0.993277	10.298191
—0.065433	—0.911797	8.370600
—0 081554	—0.129734	6.443123
—0.080329	0 152915	4.518643
—0.067468	0.336147	2.368680
—0.047361	0.419964	0.861277
—0.025874	0.404365	—1.268103
—0.007539	0 293199	—2.398731
0.004334	0.171112	—2 137685
0.010321	0 072734	—1.677440
0 012245	—0.011935	—1.217794
0.010729	—0.052895	—0.758043
0 007196	—0.080146	—0.258306
0.002869	—0.033689	0.181948
—0.001030	—0 063523	0.620769
—0.003313	—0.023903	0.750403
—0.003637	0.009634	0.549968
—0.002497	0.032634	0 349576
—0.000427	0.045696	0.149153
0 002039	0 048221	—0.031264
0 004369	0 040408	—0.251886
0 006030	0.022257	—0.412097
0.006488	—0.006231	—0.652514
0.005300	—0 035008	—0.488975
0 032617	—0.055840	—0.282961
—0.000267	—0.058945	0.082652
—0.003911	—0.047323	0.668677
—0.004836	—0.020974	0.853365
—0.004921	0.020102	0.939114
—0.002509	0.075906	1.224632
0.003162	0.143437	1.510151

Capítulo 2

UM ESTUDO POR CORRESPONDÊNCIAS DAS MICRORREGIÕES DO ESTADO DO ESPÍRITO SANTO *

2.1 Baseamo-nos para este trabalho na publicação do IBGE intitulada “Divisão do Brasil em Microrregiões Homogêneas” [12]. Nessa obra encontra-se para cada Estado da Federação uma agregação de seus municípios em unidades maiores denominadas Microrregiões. Provém tal agregação de um estudo meticoloso não só de características físico e geográficas comuns aos municípios do Estado como também do comportamento de variáveis sócio-econômicas a fim de que os novos blocos sejam da maneira mais homogênea possível “formas de organização em torno da produção” [12].

2.2 Os estudos de correspondências sobre estruturas de grafos e em especial mapas regionais têm apresentado resultados muito curiosos. No entanto tem explorado eles apenas a relação de fronteira entre as regiões do mapa [2], [15]. Ora, definindo-se um cruzamento adequado das regiões do mapa por variáveis significativas, sejam sociais, econômicas ou geográficas, deve-se obter representações bastante interessantes do mapa original nos diversos planos fatoriais. Desse modo fomos tentados a levantar para o Estado do Espírito Santo uma tabela de contingência de seus municípios por um certo conjunto de variáveis e comparar os resultados obtidos pela análise com os da agregação em Microrregiões Homogêneas. A fim de tentar reproduzir, ou pelo menos ter critérios de classificação semelhantes aos do trabalho acima mencionado, definimos as treze variáveis que se encontram relacionadas com o respectivo código:

CAFÉ	(001):	porcentagem da área destinada ao cultivo de café no município pela área total cultivada no município;
BANANA	(002):	porcentagem da área destinada ao cultivo de banana no município pela área total cultivada no município;
MILHO	(003):	porcentagem da área destinada ao cultivo de milho no município pela área total cultivada no município;
CANA	(004):	porcentagem da área destinada ao cultivo de cana no município pela área total cultivada no município;

* O arquivo para a realização desse exemplo foi gerado por Edmundo Alves Contar e Eduardo Robson Tardim Costa

LEITE	(005):	porcentagem dos litros de leite produzidos no município pela produção total de leite no Estado;
OVOS	(006):	porcentagem das dúzias de ovos de galinha produzidos no município pelo total de dúzias de ovos de galinha produzidos no Estado;
ANIMAL	(007):	valor total da produção animal do município pelo valor total da produção vegetal do município;
PRODUÇÃO	(008):	porcentagem do valor total da produção industrial no município pelo valor total da produção industrial do Estado;
EXTRATIVA	(009):	porcentagem do pessoal ocupado em indústrias extrativas de produtos minerais no município pelo total do pessoal ocupado na indústria no município,
NÃO METAL	(010):	porcentagem do número de estabelecimentos industriais de produtos de minerais não metálicos pelo número total de estabelecimentos industriais para cada município;
URBANA	(011):	porcentagem da população urbana do município sobre o total da população do município;
RENDA	(012):	porcentagem da população dos economicamente ativos de 10 anos ou mais no município que declarou perceber renda mensal igual ou superior a 1.501 cruzeiros pelo número de economicamente ativos de 10 anos ou mais que declarou perceber rendimentos mensais;
EDUCA	(013):	porcentagem do número de pessoas, do município, de 10 anos ou mais que têm o curso médio ou superior completo sobre o número de pessoas de 10 anos ou mais no município que têm o curso elementar ou médio ou superior completo.

Todas as variáveis acima foram extraídas para cada município dos dados do VIII Recenseamento Geral existentes no CI, sendo as sete primeiras oriundas do Censo Agropecuário, a oitava, nona e décima do Censo Industrial e as restantes do Demográfico. Os termos empregados em sua definição são os utilizados pelos respectivos censos e o leitor em dúvida deve recorrer às publicações próprias [5], [6], [8].

É claro que na tabela de contingência não podemos retratar explicitamente os padrões geográfico e ecológico do conjunto de municípios. O excessivo número e a forma especial das variáveis provenientes

do agropecuário tenta de certa forma suprir essa impossibilidade. As outras variáveis de cunho econômico e social tentam, entre outras coisas, separar os centros urbanos das regiões rurais. Note-se ainda que para satisfazer o princípio da homogeneidade as variáveis foram definidas de modo que todas as entradas da matriz resultem adimensionais.

Conta o Espírito Santo com 53 municípios agrupados em 8 microrregiões conforme especificado no Quadro 1. À direita de cada município e microrregião encontra-se o código utilizado nos gráficos dos resultados.

Realizou-se então uma análise de correspondências na tabela cruzada 53×13 , acima descrita, tendo-se multiplicado todas as entradas por 1.000 e truncado a parte decimal restante, a fim de se trabalhar com números inteiros.

2.3 Na Figura 1 temos os resultados numéricos gerais para a nuvem dos municípios. Um simples olhar na coluna INÉRCIA, que dá a contribuição relativa de cada ponto à inércia da nuvem (1.000 vezes aumentada) mostra-nos que os municípios 102/Montanha, 403/Domingos Martins, 505/Vitória e 804/Itapemirim são responsáveis por uma proporção bem mais elevada que os demais. Se ajuntarmos a eles os de código 103/Mucurici, 501/Cariacica e 503/Viana, teremos grande parte da inércia total explicada por esse conjunto.

O Município de Vitória explica-se por si só como um ponto bastante atípico dentro da nuvem. Montanha apresenta também uma relação curiosa entre os valores das variáveis MILHO(3), OVOS(6), ANIMAL(7), URBANA(11) pois todos eles se apresentam elevados, porém as três primeiras são essencialmente rurais enquanto que a última indica uma alta densidade urbana no município (vide Fig. 3). Domingos Martins também se revela um caso mais especial já dentro de uma microrregião como é a 4. Na realidade comparando-se os valores das variáveis agrícolas nesse município com os no município apontado como mais agrícola da região [12] nota-se que as variáveis escolhidas ressaltam tanto ou mais Domingos Martins do que Afonso Cláudio:

	Afonso Cláudio	Domingos Martins
1	28	12
2	27	397
3	17	139
4	305	236
5	25	6
6	651	618
7	307	1.524

esses pontos podem na saída gráfica apresentar posições discrepantes com relação à totalidade da nuvem. Da mesma forma para as variáveis coluna, os pontos 003 e 005 apresentam valores bem elevados para a INÉRCIA em relação a seus companheiros (Fig. 2). Passemos então à análise do plano dos dois primeiros eixos fatoriais.

ESPÍRITO SANTO

QUADRO 1

MICRORREGIÕES HOMOGÊNEAS DO ESTADO DO ESPÍRITO SANTO COM OS RESPECTIVOS MUNICÍPIOS

Microrregiões Homogêneas

- | | |
|-----------------------------|------------------------------|
| (1) <i>Microrregião 203</i> | (5) <i>Microrregião 207</i> |
| 101. Ecoporanga | 501. Cariacica |
| 102. Montanha | 502. Serra |
| 103. Mucurici | 503. Viana |
| | 504. Vila Velha |
| | 505. Vitória |
| (2) <i>Microrregião 204</i> | (6) <i>Microrregião 208</i> |
| 201. Baixo Guandu | 601. Castelo |
| 202. Barra de São Francisco | 602. Conceição do Castelo |
| 203. Boa Esperança | 603. Iúna |
| 204. Colatina | 604. Muniz Freire |
| 205. Mantenópolis | |
| 206. Nova Venécia | (7) <i>Microrregião 209</i> |
| 207. Pancas | 701. Alegre |
| 208. São Gabriel da Palha | 702. Apiacá |
| | 703. Atilio Vivacqua |
| (3) <i>Microrregião 205</i> | 704. Bom Jesus do Norte |
| 301. Aracruz | 705. Cachoeiro do Itapemirim |
| 302. Conceição da Barra | 706. Divino de São Lourenço |
| 303. Fundão | 707. Dolores do Rio Preto |
| 304. Linhares | 708. Guaçuí |
| 305. Pinheiros | 709. Jerônimo Monteiro |
| 306. São Mateus | 710. Mimoso do Sul |
| | 711. Muqui |
| | 712. São José do Calçado |
| (4) <i>Microrregião 206</i> | (8) <i>Microrregião 210</i> |
| 401. Afonso Cláudio | 801. Anchieta |
| 402. Alfredo Chaves | 802. Guarapari |
| 403. Domingos Martins | 803. Iconha |
| 404. Ibiraja | 804. Itapemirim |
| 405. Itaguaçu | 805. Piúma |
| 406. Itarana | 806. Presidente Kennedy |
| 407. Santa Leopoldina | 807. Rio Novo do Sul |
| 408. Santa Tereza | |

FIGURA 1

Resultados numéricos gerais para a nuvem dos municípios

***** PAGINA 112

***** ANALISE DE CORRESPONDENCIA *****

QUALIDADE DA REPRESENTACAO PARA O CONJUNTO DAS LINHAS							
IVARIABELI	INUMEROI	DESCRIPC	PESO	VARIANCIA	INERCIA	Q.L.T.	
ESANTO	0	101	14.820691	323.617920	3.545521	0.006121	
	0	102	74.230362	7395.062500	81.019455	0.007820	
	0	103	43.351303	4531.679688	49.648560	0.007524	
	0	201	16.050446	568.879883	6.232581	0.004822	
	0	202	12.392889	726.819336	7.962949	0.007512	
	0	203	14.487464	541.693359	5.934730	0.006595	
	0	204	16.756577	982.258057	10.761506	0.005698	
	0	205	11.258329	914.611816	10.020382	0.006018	
	0	206	15.288795	403.849121	4.424523	0.006208	
	0	207	11.131383	1132.505127	12.407597	0.007426	
	0	208	12.496032	950.967529	10.418691	0.007003	
	0	301	15.685494	200.168289	2.193021	0.006439	
	0	302	21.723251	321.652832	3.523992	0.006852	
	0	303	16.328140	561.080566	6.147135	0.005249	
	0	304	12.972072	1142.522949	12.517351	0.006349	
	0	305	22.302429	382.897705	4.194982	0.005372	
	0	306	14.709614	315.505859	3.456646	0.007295	
	0	401	12.543633	766.233643	8.394766	0.005034	
	0	402	13.789269	277.843994	3.044027	0.003468	
	0	403	24.873047	6290.746094	68.920670	0.007717	
	0	404	20.461746	198.195526	2.171407	0.005673	
	0	405	15.320532	241.314331	2.643812	0.006764	
	0	406	14.170102	454.636963	4.980947	0.006653	
	0	407	15.590287	1127.968262	12.357890	0.007218	
	0	408	14.209773	349.478516	3.828847	0.006828	
	0	501	20.096786	4774.050781	52.303955	0.006971	
	0	502	15.915580	855.020752	9.367510	0.007140	
	0	503	16.201202	3979.379639	43.597626	0.005801	
	0	504	26.055206	1812.838379	19.861237	0.006969	
	0	505	22.397644	6442.773438	70.586258	0.007603	
	0	601	14.527130	475.206787	5.206308	0.007614	
	0	602	14.170102	1497.376465	16.405075	0.006869	
	0	603	11.583622	935.960693	10.254278	0.007790	
	0	604	12.496032	429.876953	4.709682	0.006904	
	0	701	17.113602	465.883301	5.104161	0.005243	
	0	702	26.142487	785.697021	8.608006	0.007578	
	0	703	16.225006	230.162949	2.521639	0.003397	
	0	704	30.157089	595.678467	6.526183	0.006168	
	0	705	22.334167	1234.857910	13.528961	0.004557	
	0	706	12.789588	1052.992188	11.536460	0.005736	
	0	707	12.662645	554.800537	6.078328	0.006949	
	0	708	21.826401	677.649414	7.424250	0.004155	
	0	709	17.803864	621.517578	6.809275	0.002760	
	0	710	19.739761	156.514343	1.714753	0.001403	

***** ANALISE DE CORRESPONDENCIA *****

QUALIDADE DA REPRESENTACAO PARA O CONJUNTO DAS LINHAS								
IVARIABELI	CATEGORIA	INUMEROI	DESCRICA	PESO	VARIANCIA	INERCIA	Q.L.T.	
COLUNA	0	711		19.366852	276.696289	3.031453	0.007470	
	0	712		22.556320	311.433350	3.412028	0.006820	
	0	801		13.900347	1899.884033	20.814911	0.005552	
	0	802		14.558867	4480.699219	49.090027	0.006676	
	0	803		14.558867	1099.594482	12.047029	0.004502	
	0	804		17.367493	15864.859375	173.813568	0.007869	
	0	805		31.489990	1071.669678	11.741090	0.006569	
	0	806		22.405579	1334.095703	14.616199	0.004412	
	0	807		16.613770	338.501953	3.708589	0.004340	

FIGURA, 2

Resultados numéricos gerais para a nuvem das variáveis

***** PAGINA 114

***** ANÁLISE DE CORRESPONDÊNCIA *****

QUALIDADE DA REPRESENTAÇÃO PARA O CONJUNTO DAS COLUNAS							
INOME DA VARIÁVEL	CATEGORIA	DESCRICAÇÃO	PESO	VARIÂNCIA	INÉRCIA	Q.L.T.	
COLUNA	0	001	7.719769	897.801025	9.836204	0.002707	
	0	002	7.767371	7838.656250	85.879379	0.007676	
	0	003	57.235794	117842.109375	195.476105	0.007425	
	0	004	159.314453	10300.976563	112.856323	0.006949	
	0	005	23.794022	17196.054688	188.397995	0.007880	
	0	006	179.347809	5410.769531	59.279770	0.005090	
	0	007	375.618652	13157.851563	144.155914	0.007891	
	0	008	7.735636	4687.398438	51.354599	0.004408	
	0	011	139.844467	5157.406250	56.503952	0.006195	
	0	012	1.872420	769.785156	8.433677	0.002912	
	0	013	31.426529	1049.870117	11.502258	0.004387	
	0	009	1.420183	702.331299	7.694661	0.002619	
	0	010	6.902568	6264.601563	68.634201	0.006081	

PROJEÇÕES DOS PONTOS

FIGURA 3

A matriz de contigência analisada. Na primeira coluna tem-se a entrada correspondente ao município da quarta coluna e a variável da sexta

TABELA DE CONTINGÊNCIA

IN=	1	FIM=	578
CONTAGEM	CÉLULA	***	
36	0	2	101
9	0	2	101
24	0	2	101
428	0	2	101
52	0	2	101
494	0	2	101
631	0	2	101
3	0	2	101
140	0	2	101
1	0	2	101
50	0	2	101
15	0	2	102
304	0	2	102
167	0	2	102
18	0	2	102
510	0	2	102
7436	0	2	102
40	0	2	102
96	0	2	102
706	0	2	102
3	0	2	102
61	0	2	102
41	0	2	103
1	0	2	103
11	0	2	103
60	0	2	103
7	0	2	103
920	0	2	103
4315	0	2	103
1	0	2	103
67	0	2	103
3	0	2	103
38	0	2	103
17	0	2	201
6	0	2	201
32	0	2	201
345	0	2	201
18	0	2	201
602	0	2	201
422	0	2	201
4	0	2	201
2	0	2	201
9	0	2	201
486	0	2	201
4	0	2	201
76	0	2	201
16	0	2	202
10	0	2	202
27	0	2	202
453	0	2	202
18	0	2	202
500	0	2	202
162	0	2	202
7	0	2	202
21	0	2	202
30	0	2	202
257	0	2	202
2	0	2	202

TABELA DE CONTINGÊNCIA

IN=	1	FIM=	578
CONTAGEM	CÉLULA	***	
59	0	2	202
10	0	2	203
3	0	2	203
6	0	2	203
560	0	2	203
6	0	2	203
426	0	2	203
673	0	2	203
1	0	2	203
104	0	2	203
37	0	2	203
62	0	2	204
32	0	2	204
17	0	2	204
570	0	2	204
23	0	2	204
389	0	2	204
330	0	2	204
63	0	2	204
5	0	2	204
31	0	2	204
502	0	2	204
8	0	2	204
80	0	2	204
2	0	2	205
1	0	2	205
18	0	2	205
393	0	2	205
21	0	2	205
566	0	2	205
141	0	2	205
243	0	2	205
8	0	2	205
26	0	2	205
53	0	2	206
17	0	2	206
26	0	2	206
496	0	2	206
26	0	2	206
451	0	2	206
517	0	2	206
19	0	2	206
5	0	2	206
15	0	2	206
228	0	2	206
5	0	2	206
69	0	2	206
14	0	2	207
3	0	2	207
10	0	2	207
613	0	2	207
3	0	2	207
372	0	2	207
153	0	2	207
5	0	2	207
163	0	2	207
67	0	2	207
17	0	2	208

(continua)

TABELA DE CONTINGÊNCIA

IN=	1	FIM=	578
CONTAGEM	CÉLULA	***	
12	0	2 208	3
36	0	2 208	3
630	0	2 208	3
19	0	2 208	3
314	0	2 208	3
180	0	2 208	3
20	0	2 208	3
295	0	2 208	3
1	0	2 208	3
51	0	2 208	3
8	0	2 301	3
4	0	2 301	3
49	0	2 301	3
457	0	2 301	3
74	0	2 301	3
418	0	2 301	3
613	0	2 301	3
4	0	2 301	3
309	0	2 301	3
1	0	2 301	3
40	0	2 301	3
27	0	2 302	3
8	0	2 302	3
75	0	2 302	3
383	0	2 302	3
38	0	2 302	3
502	0	2 302	3
1411	0	2 302	3
16	0	2 302	3
212	0	2 302	3
1	0	2 302	3
65	0	2 302	3
5	0	2 303	3
1	0	2 303	3
162	0	2 303	3
537	0	2 303	3
21	0	2 303	3
278	0	2 303	3
451	0	2 303	3
2	0	2 303	3
465	0	2 303	3
1	0	2 303	3
135	0	2 303	3
40	0	2 304	3
17	0	2 304	3
48	0	2 304	3
653	0	2 304	3
12	0	2 304	3
284	0	2 304	3
165	0	2 304	3
44	0	2 304	3
303	0	2 304	3
2	0	2 304	3
67	0	2 304	3
8	0	2 305	3
11	0	2 305	3
347	0	2 305	3
21	0	2 305	3
620	0	2 305	3
1260	0	2 305	3
2	0	2 305	3
496	0	2 305	3
3	0	2 305	3
43	0	2 305	3

(continuação)

TABELA DE CONTINGÊNCIA

IN=	1	FIM=	578
CONTAGEM	CÉLULA	***	
14	0	2 306	3
5	0	2 306	3
36	0	2 306	3
426	0	2 306	3
57	0	2 306	3
479	0	2 306	3
433	0	2 306	3
8	0	2 306	3
311	0	2 306	3
5	0	2 306	3
80	0	2 306	3
28	0	2 401	3
27	0	2 401	3
17	0	2 401	3
305	0	2 401	3
25	0	2 401	3
651	0	2 401	3
307	0	2 401	3
6	0	2 401	3
5	0	2 401	3
11	0	2 401	3
155	0	2 401	3
1	0	2 401	3
43	0	2 401	3
8	0	2 402	3
11	0	2 402	3
217	0	2 402	3
314	0	2 402	3
43	0	2 402	3
424	0	2 402	3
456	0	2 402	3
2	0	2 402	3
214	0	2 402	3
49	0	2 402	3
12	0	2 403	3
397	0	2 403	3
139	0	2 403	3
236	0	2 403	3
6	0	2 403	3
618	0	2 403	3
1524	0	2 403	3
3	0	2 403	3
149	0	2 403	3
2	0	2 403	3
49	0	2 403	3
13	0	2 404	3
5	0	2 404	3
36	0	2 404	3
508	0	2 404	3
56	0	2 404	3
398	0	2 404	3
1044	0	2 404	3
14	0	2 404	3
454	0	2 404	3
4	0	2 404	3
47	0	2 404	3
15	0	2 405	3
10	0	2 405	3
36	0	2 405	3
412	0	2 405	3
24	0	2 405	3
525	0	2 405	3
588	0	2 405	3
4	0	2 405	3

(continua)

(continuação)

TABELA DE CONTINGÊNCIA

IN=	1	FIM=	578
CONTAGEM	CÉLULA	***	
252	0 2	405 3	11
4	0 2	405 3	12
61	0 2	405 3	13
5	0 2	406 3	1
7	0 2	406 3	2
2	0 2	406 3	3
429	0 2	406 3	4
12	0 2	406 3	5
554	0 2	406 3	6
501	0 2	406 3	7
207	0 2	406 3	11
2	0 2	406 3	12
67	0 2	406 3	13
11	0 2	407 3	1
114	0 2	407 3	2
76	0 2	407 3	3
266	0 2	407 3	4
10	0 2	407 3	5
647	0 2	407 3	6
685	0 2	407 3	7
1	0 2	407 3	8
15	0 2	407 3	9
25	0 2	407 3	10
69	0 2	407 3	11
1	0 2	407 3	12
45	0 2	407 3	13
17	0 2	408 3	1
25	0 2	408 3	2
17	0 2	408 3	3
405	0 2	408 3	4
63	0 2	408 3	5
513	0 2	408 3	6
496	0 2	408 3	7
10	0 2	408 3	8
2	0 2	408 3	9
12	0 2	408 3	10
163	0 2	408 3	11
4	0 2	408 3	12
64	0 2	408 3	13
2	0 2	501 3	1
10	0 2	501 3	2
671	0 2	501 3	3
97	0 2	501 3	4
110	0 2	501 3	5
120	0 2	501 3	6
550	0 2	501 3	7
209	0 2	501 3	8
682	0 2	501 3	11
6	0 2	501 3	12
76	0 2	501 3	13
7	0 2	502 3	1
7	0 2	502 3	2
338	0 2	502 3	3
357	0 2	502 3	4
79	0 2	502 3	5
223	0 2	502 3	6
413	0 2	502 3	7
42	0 2	502 3	8
460	0 2	502 3	11
2	0 2	502 3	12
78	0 2	502 3	13
9	0 2	503 3	1
9	0 2	503 3	2
754	0 2	503 3	3

TABELA DE CONTINGÊNCIA

IN=	1	FIM=	578
CONTAGEM	CÉLULA	***	
63	0 2	503 3	4
27	0 2	503 3	5
154	0 2	503 3	6
566	0 2	503 3	7
1	0 2	503 3	8
15	0 2	503 3	9
250	0 2	503 3	10
153	0 2	503 3	11
1	0 2	503 3	12
40	0 2	503 3	13
3	0 2	504 3	1
8	0 2	504 3	2
344	0 2	504 3	3
215	0 2	504 3	5
440	0 2	504 3	6
1024	0 2	504 3	7
89	0 2	504 3	8
3	0 2	504 3	9
36	0 2	504 3	10
984	0 2	504 3	11
19	0 2	504 3	12
119	0 2	504 3	13
787	0 2	505 3	3
24	0 2	505 3	5
187	0 2	505 3	6
345	0 2	505 3	7
186	0 2	505 3	8
1	0 2	505 3	9
18	0 2	505 3	10
992	0 2	505 3	11
53	0 2	505 3	12
230	0 2	505 3	13
21	0 2	601 3	1
8	0 2	601 3	2
49	0 2	601 3	3
494	0 2	601 3	4
36	0 2	601 3	5
419	0 2	601 3	6
307	0 2	601 3	7
9	0 2	601 3	8
14	0 2	601 3	9
49	0 2	601 3	10
329	0 2	601 3	11
9	0 2	601 3	12
87	0 2	601 3	13
6	0 2	602 3	1
127	0 2	602 3	2
71	0 2	602 3	3
456	0 2	602 3	4
2	0 2	602 3	5
469	0 2	602 3	6
441	0 2	602 3	7
2	0 2	602 3	8
151	0 2	602 3	11
29	0 2	602 3	12
32	0 2	602 3	13
14	0 2	603 3	1
3	0 2	603 3	2
4	0 2	603 3	3
530	0 2	603 3	4
9	0 2	603 3	5
455	0 2	603 3	6
158	0 2	603 3	7
1	0 2	603 3	8
214	0 2	603 3	11

(continua)

(continuação)

TABELA DE CONTINGÊNCIA

IN=	1	FIM=	578		
CONTAGEM	CÉLULA	***			
3	0	2	603	3	12
69	0	2	603	3	13
18	0	2	604	3	1
5	0	2	604	3	2
55	0	2	604	3	3
447	0	2	604	3	4
39	0	2	604	3	5
456	0	2	604	3	6
329	0	2	604	3	7
2	0	2	604	3	8
171	0	2	604	3	11
3	0	2	604	3	12
50	0	2	604	3	13
58	0	2	701	3	1
11	0	2	701	3	2
19	0	2	701	3	3
494	0	2	701	3	4
6	0	2	701	3	5
479	0	2	701	3	6
638	0	2	701	3	7
3	0	2	701	3	8
303	0	2	701	3	11
5	0	2	701	3	12
141	0	2	701	3	13
9	0	2	702	3	1
1	0	2	702	3	2
11	0	2	702	3	3
360	0	2	702	3	4
205	0	2	702	3	5
422	0	2	702	3	6
1757	0	2	702	3	7
408	0	2	702	3	11
4	0	2	702	3	12
118	0	2	702	3	13
13	0	2	703	3	1
2	0	2	703	3	2
140	0	2	703	3	3
350	0	2	703	3	4
5	0	2	703	3	5
503	0	2	703	3	6
846	0	2	703	3	7
157	0	2	703	3	11
1	0	2	703	3	12
28	0	2	703	3	13
7	0	2	704	3	1
142	0	2	704	3	3
320	0	2	704	3	4
15	0	2	704	3	5
521	0	2	704	3	6
1880	0	2	704	3	7
8	0	2	704	3	8
737	0	2	704	3	11
6	0	2	704	3	12
165	0	2	704	3	13
70	0	2	705	3	1
19	0	2	705	3	2
72	0	2	705	3	3
354	0	2	705	3	4
134	0	2	705	3	5
439	0	2	705	3	6
625	0	2	705	3	7
138	0	2	705	3	8
18	0	2	705	3	9
179	0	2	705	3	10
630	0	2	705	3	11

TABELA DE CONTINGÊNCIA

IN=	1	FIM=	578		
CONTAGEM	CÉLULA	***			
8	0	2	705	3	12
129	0	2	705	3	13
6	0	2	706	3	1
701	0	2	706	3	4
298	0	2	706	3	6
482	0	2	706	3	7
95	0	2	706	3	11
30	0	2	706	3	13
9	0	2	707	3	1
15	0	2	707	3	3
533	0	2	707	3	4
45	0	2	707	3	5
405	0	2	707	3	6
397	0	2	707	3	7
153	0	2	707	3	11
39	0	2	707	3	13
42	0	2	708	3	1
3	0	2	708	3	2
8	0	2	708	3	3
734	0	2	708	3	4
6	0	2	708	3	5
250	0	2	708	3	6
1004	0	2	708	3	7
13	0	2	708	3	8
560	0	2	708	3	11
4	0	2	708	3	12
127	0	2	708	3	13
8	0	2	709	3	1
1	0	2	709	3	2
5	0	2	709	3	3
223	0	2	709	3	4
770	0	2	709	3	6
827	0	2	709	3	7
340	0	2	709	3	11
70	0	2	709	3	13
40	0	2	710	3	1
15	0	2	710	3	2
223	0	2	710	3	3
349	0	2	710	3	4
54	0	2	710	3	5
372	0	2	710	3	6
897	0	2	710	3	7
5	0	2	710	3	8
27	0	2	710	3	9
75	0	2	710	3	10
305	0	2	710	3	11
1	0	2	710	3	12
125	0	2	710	3	13
18	0	2	711	3	1
3	0	2	711	3	2
4	0	2	711	3	3
483	0	2	711	3	4
2	0	2	711	3	5
509	0	2	711	3	6
955	0	2	711	3	7
2	0	2	711	4	8
367	0	2	711	3	11
4	0	2	711	3	12
94	0	2	711	3	13
13	0	2	712	3	1
2	0	2	712	3	2
58	0	2	712	3	3
528	0	2	712	3	4
9	0	2	712	3	5
403	0	2	712	3	6

(continua)

(conclusão)

TABELA DE CONTINGÊNCIA

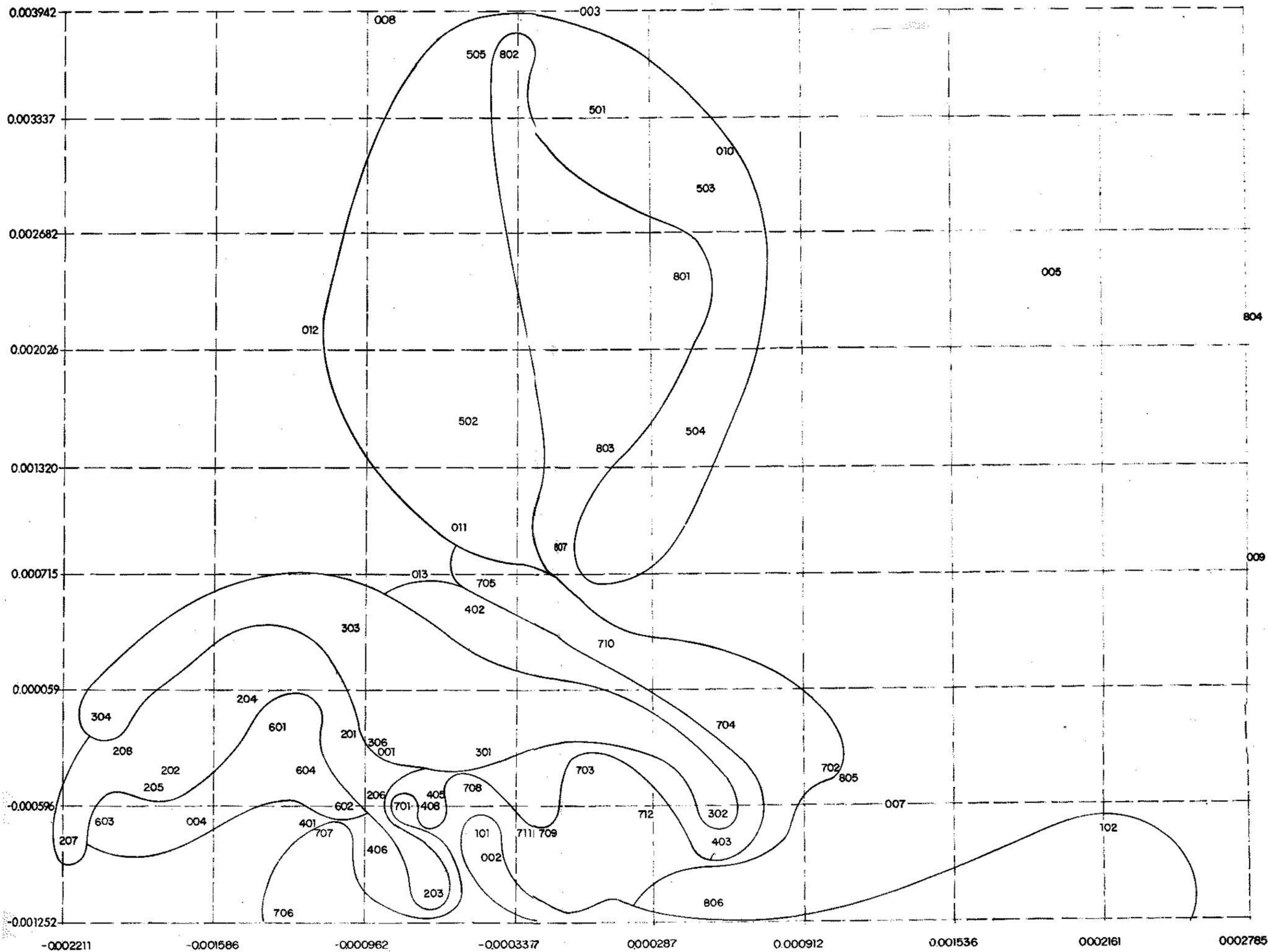
IN=	1	FIM=	578		
CONTAGEM	CÉLULA	***			
1367	0	2	712	3	7
372	0	2	712	3	11
91	0	2	712	3	13
7	0	2	801	3	1
3	0	2	801	3	2
466	0	2	801	3	3
136	0	2	801	3	4
161	0	2	801	3	5
235	0	2	801	3	6
454	0	2	801	3	7
199	0	2	801	3	11
3	0	2	801	3	12
88	0	2	801	3	13
4	0	2	802	3	1
7	0	2	802	3	2
731	0	2	802	3	3
98	0	2	802	3	4
7	0	2	802	3	5
163	0	2	802	3	6
253	0	2	802	3	7
3	0	2	802	3	8
4	0	2	802	3	9
467	0	2	802	3	11
5	0	2	802	3	12
93	0	2	802	3	13
4	0	2	803	3	1
2	0	2	803	3	2
426	0	2	803	3	3
300	0	2	803	3	4
44	0	2	803	3	5
228	0	2	803	3	6
606	0	2	803	3	7
176	0	2	803	3	11
49	0	2	803	3	13
42	0	2	804	3	1
6	0	2	804	3	2
14	0	2	804	3	3

TABELA DE CONTINGÊNCIA

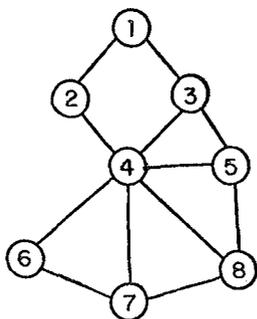
IN=	1	FIM=	578		
CONTAGEM	CÉLULA	***			
3	0	2	804	3	4
937	0	2	804	3	5
44	0	2	804	3	6
757	0	2	804	3	7
22	0	2	804	3	8
2	0	2	804	3	9
34	0	2	804	3	10
258	0	2	804	3	11
1	0	2	804	3	12
69	0	2	804	3	13
88	0	2	805	3	3
132	0	2	805	3	4
49	0	2	805	3	5
729	0	2	805	3	6
2268	0	2	805	3	7
629	0	2	805	3	11
74	0	2	805	3	13
34	0	2	806	3	1
1	0	2	806	3	2
11	0	2	806	3	3
779	0	2	806	3	4
29	0	2	806	3	5
179	0	2	806	3	6
1656	0	2	806	3	7
1	0	2	806	3	8
44	0	2	806	3	11
1	0	2	806	3	12
89	0	2	806	3	13
15	0	2	807	3	1
1	0	2	807	3	2
276	0	2	807	3	3
366	0	2	807	3	4
77	0	2	807	3	5
278	0	2	807	3	6
647	0	2	807	3	7
340	0	2	807	3	11
2	0	2	807	3	12
92	0	2	807	3	13

Saída no primeiro plano fatorial

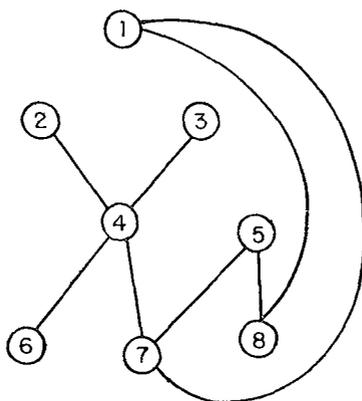
ANÁLISE DE CORRESPONDÊNCIA



2.4 Nas Figuras 4, 5, 6 temos respectivamente a saída gráfica e as saídas numéricas para as duas nuvens. Concentra-se nesse plano 53% da inércia total (= 0,723376), valendo o primeiro autovalor 0,210281 e o segundo 0,170536. Tomando-se o mapa do Espírito Santo (Fig. 7) o grafo obtido através a relação “faz fronteira com” para as microrregiões, tem a seguinte forma:



da saída gráfica podemos obter o seguinte grafo:



A valência de cada vértice do segundo grafo é igual ou menor que a do respectivo no primeiro. Isso é evidente pois apenas parte da riqueza da estrutura foi retratada no cruzamento. A microrregião 4 é a que apresenta mais ligações com as demais, tendo esse fato sido preservado. Em particular por ter ela características comuns com todas as regiões fronteiriças vemos que na Figura 4 são os seus pontos os que se apresentam mais dispersos pelo aglomerado do canto inferior esquerdo. As micro 5 e 8, conforme já era mais ou menos previsto pela análise anterior de algum de seus municípios foram extremamente separadas pelo método, a notar inclusive o visível rompimento de 5 e 8 com 4. Ora, pode-se dizer que tanto 5 com suas características essenciais urbanas e industriais,

Resultados numéricos para o primeiro eixo fatorial

***** ANALISE DE CORRESPONDENCIA *****

RESULTADOS FATORIAIS PARA O CONJUNTO DAS LINHAS							
NOME DA VARIÁVEL	CATEGORIA		FATOR				
	INUMERO	DESCRICA	PROJECAO	CORRELACAO	CTR. RELATIVA		
ESANTO	0	101	-0.000782	0.003532	0.000043		
	0	102	-0.000774	0.000758	0.000211		
	0	103	-0.001302	0.002045	0.000350		
	0	201	-0.000258	0.000237	0.000005		
	0	202	-0.000453	0.000442	0.000012		
	0	203	-0.001161	0.004546	0.000093		
	0	204	-0.000055	0.000006	0.000000		
	0	205	-0.000570	0.000504	0.000017		
	0	206	-0.000596	0.001695	0.000026		
	0	207	-0.000851	0.000897	0.000038		
	0	208	-0.000335	0.000186	0.000007		
	0	301	-0.000351	0.001215	0.000009		
	0	302	-0.000675	0.003878	0.000047		
	0	303	0.000352	0.000455	0.000010		
	0	304	-0.000134	0.000026	0.000001		
	0	305	-0.000685	0.003440	0.000050		
	0	306	-0.000305	0.000547	0.000007		
	0	401	-0.000761	0.001195	0.000035		
	0	402	0.000477	0.001424	0.000015		
	0	403	-0.000828	0.000342	0.000081		
	0	404	-0.000455	0.002698	0.000020		
	0	405	-0.000615	0.003026	0.000028		
	0	406	-0.000894	0.003140	0.000054		
	0	407	-0.000822	0.001177	0.000050		
	0	408	-0.000639	0.002094	0.000028		
	0	501	0.003337	0.005908	0.001064		
	0	502	0.001595	0.005971	0.000193		
	0	503	0.002869	0.004224	0.000634		
	0	504	0.001510	0.004131	0.000283		
	0	505	0.003625	0.005759	0.001400		
	0	601	-0.000194	0.000146	0.000003		
	0	602	-0.000658	0.000517	0.000029		
	0	603	-0.000763	0.000907	0.000032		
	0	604	-0.000466	0.000796	0.000013		
	0	701	-0.000645	0.001926	0.000034		
	0	702	-0.000455	0.000870	0.000026		
	0	703	-0.000433	0.001667	0.000014		
	0	704	-0.000212	0.000287	0.000006		
	0	705	0.000633	0.000913	0.000043		
	0	706	-0.001254	0.002406	0.000096		
	0	707	-0.000802	0.001852	0.000039		
	0	708	-0.000542	0.001194	0.000031		
	0	709	-0.000792	0.002262	0.000053		
	0	710	0.000248	0.000981	0.000006		

***** ANALISE DE CORRESPONDENCIA *****

RESULTADOS FATORIAIS PARA O CONJUNTO DAS LINHAS						
NOME DA VARIAVEL	CATEGORIA		FATOR			
	INUMERO	DESCRICA	PROJECAO	CORRELACAO	CTR. RELATIVA	
	0	711	-0.000786I	0.005450I	0.000057I	
	0	712	-0.000686I	0.004300I	0.000051I	
	0	801	0.002385I	0.005247I	0.000376I	
	0	802	0.003633I	0.005404I	0.000914I	
	0	803	0.001439I	0.003456I	0.000143I	
	0	804	0.002107I	0.000613I	0.000367I	
	0	805	-0.000515I	0.000983I	0.000040I	
	0	806	-0.001211I	0.003104I	0.000156I	
	0	807	0.000800I	0.003963I	0.000051I	

***** ANALISE DE CORRESPONDENCIA *****

RESULTADOS FATORIAIS PARA O CONJUNTO DAS COLUNAS						
NOME DA VARIAVEL	CATEGORIA		FATOR			
	INUMERO	DESCRICA	PROJECAO	CORRELACAO	CTR. RELATIVA	
COLUNA	0	001	-0.000320I	0.000111I	0.000004I	
	0	002	-0.000963I	0.000116I	0.000034I	
	0	003	0.003942I	0.006284I	0.004230I	
	0	004	-0.000761I	0.001128I	0.000439I	
	0	005	0.002382I	0.000989I	0.000642I	
	0	006	-0.000625I	0.001634I	0.000334I	
	0	007	-0.000645I	0.001497I	0.000743I	
	0	008	0.003839I	0.003066I	0.000542I	
	0	011	0.000906I	0.002804I	0.000546I	
	0	012	0.002052I	0.001291I	0.000037I	
	0	013	0.000658I	0.001633I	0.000065I	
	0	009	0.000759I	0.000147I	0.000004I	
	0	010	0.003095I	0.001330I	0.000314I	

FIGURA 6

Resultados numéricos para o segundo eixo fatorial

***** PAGINA
 ***** ANALISE DE CORRESPONDENCIA *****

RESULTADOS FATORIAIS PARA O CONJUNTO DAS LINHAS							
NOME DA VARIÁVEL	CATEGORIA		FATOR				
	INUMERO	DESCRICA	PROJECAO	CORRELACAO	CTR. RELATIVA		
ESANTO	0	101	-0.000472	0.001287	0.000019		
	0	102	0.002182	0.006025	0.002073		
	0	103	0.002027	0.004954	0.001044		
	0	201	-0.001023	0.003722	0.000099		
	0	202	-0.001768	0.006716	0.000227		
	0	203	-0.000667	0.001498	0.000038		
	0	204	-0.001459	0.004577	0.000209		
	0	205	-0.001828	0.005184	0.000221		
	0	206	-0.000927	0.004104	0.000077		
	0	207	-0.002211	0.006054	0.000319		
	0	208	-0.001962	0.006374	0.000282		
	0	301	-0.000476	0.002239	0.000021		
	0	302	0.000552	0.002589	0.000039		
	0	303	-0.001018	0.003798	0.000099		
	0	304	-0.002047	0.005998	0.000319		
	0	305	0.000208	0.000318	0.000006		
	0	306	-0.000914	0.004904	0.000072		
	0	401	-0.001181	0.002880	0.000103		
	0	402	-0.000518	0.001679	0.000022		
	0	403	0.000584	0.000170	0.000050		
	0	404	-0.000039	0.000019	0.000000		
	0	405	-0.000667	0.003563	0.000040		
	0	406	-0.000928	0.003380	0.000071		
	0	407	-0.000268	0.000125	0.000007		
	0	408	-0.000701	0.002516	0.000041		
	0	501	0.000032	0.000001	0.000000		
	0	502	-0.000543	0.000691	0.000027		
	0	503	0.000510	0.000134	0.000025		
	0	504	0.000459	0.000382	0.000032		
	0	505	-0.000525	0.000121	0.000036		
	0	601	-0.001311	0.006627	0.000147		
	0	602	-0.001058	0.001334	0.000093		
	0	603	-0.002034	0.006455	0.000281		
	0	604	-0.001196	0.005242	0.000105		
	0	701	-0.000815	0.003073	0.000067		
	0	702	0.001015	0.004321	0.000158		
	0	703	-0.000030	0.000008	0.000000		
	0	704	0.000588	0.002206	0.000061		
	0	705	-0.000463	0.000489	0.000028		
	0	706	-0.001303	0.002600	0.000127		
	0	707	-0.001129	0.003668	0.000095		
	0	708	-0.000511	0.001059	0.000033		
	0	709	-0.000224	0.000181	0.000005		
	0	710	0.000060	0.000058	0.000000		

como 8 com sua baixa atividade agropastoril e uma economia tendendo à urbanização, com parte da renda advinda inclusive da exploração turística (Guarapari, etc.), são, exceto geograficamente, bem divorciadas de 4 que ocupa posição relativamente central no conjunto das micro essencialmente agrícolas, a saber 2, 3 e 6. Mais interessante ainda é ver que, assim como 4 se apresenta dispersa no aglomerado das agrícolas, 7 também se espalha por esse conjunto mas já apresenta maior proximidade com 5 e 8, chegando inclusive a manter fronteira com elas. Essa região é a que apresenta justamente de forma significativa propriedades comuns aos dois grupos anteriores com uma forte atividade agropastoril paralelamente a uma crescente industrialização e urbanização, conforme atestam os dados de [6] e [8].

Com a microrregião 1 o fato é mais curioso. Na verdade os pontos 102 e 103 encontram-se nas margens do círculo da guarda para 95% de confiança (Quadro 2), devendo inclusive ser retirados em uma segunda análise. Aliás o ponto 102, pela análise da Figura 1 já apresentava características peculiares. Não possui ainda essa região um caráter tão próspero na agropecuária como as demais, embora seja zona de grande potencial de desenvolvimento. Devido mesmo a essa promessa, para lá afluem migrantes do sul do Estado, bem como de Minas Gerais e do nordeste sendo ela também semelhante aos municípios 302 e 305 [12]. Essa relação foi em parte preservada pela estrutura dos pontos 101, 302 e 305 (o ponto 302 não foi impresso por ser coincidente com o 711).

Quanto às variáveis classificatórias, embora tenham se apresentado dispersas pelo gráfico, quase todas se encontram dentro dos respectivos círculos de guarda a 95% de confiança (Quadro 2). As de número 1, 2 e 4 são as que mais infiltradas estão no aglomerado agrícola. Ora, são elas relativas ao café, banana e cana-de-açúcar que são exatamente os produtos vegetais mais típicos do Estado. O café, que é cultivado em todas as microrregiões do Espírito Santo está numa posição bem central com respeito à nuvem global dos municípios. As variáveis demográficas 13 (EDUCA) e especialmente 11 (URBANA) e a variável 9 (EXTRATIVA) aparecem como pontos de equilíbrio entre a massa rural e a massa urbana. São realmente centro de simetria entre os pontos da metade superior e a inferior. URBANA opõe muito nitidamente o par Vitória-Guarapari (505-802) e mais Cariacica (501), situados no extremo superior, dos municípios rurais no extremo inferior. As de código 008, 010 e 012 por um lado e 007 por outro acentuam a separação mencionada, enquanto que 005 está fora do círculo de guarda.

Com a variável MILHO (003) se nos depara uma situação curiosa. A sua projeção no 1.º eixo é bastante estranha, encontrando-se ela próxima dos centros urbanos e distanciada dos municípios nos quais o seu valor devia ser elevado. Tal fato pode indicar um erro no levantamento

O plano 1x3

ANÁLISE DE CORRESPONDÊNCIA

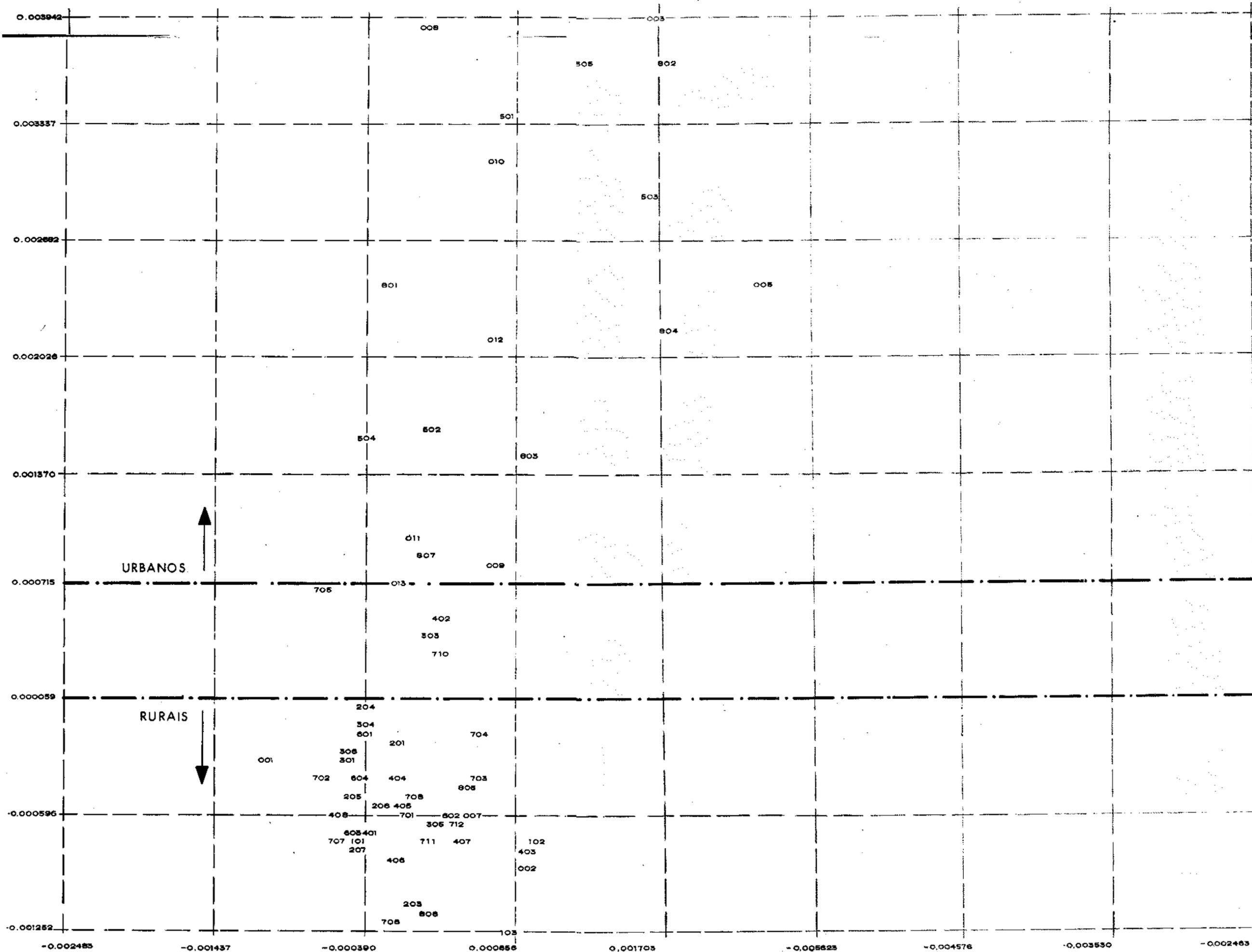
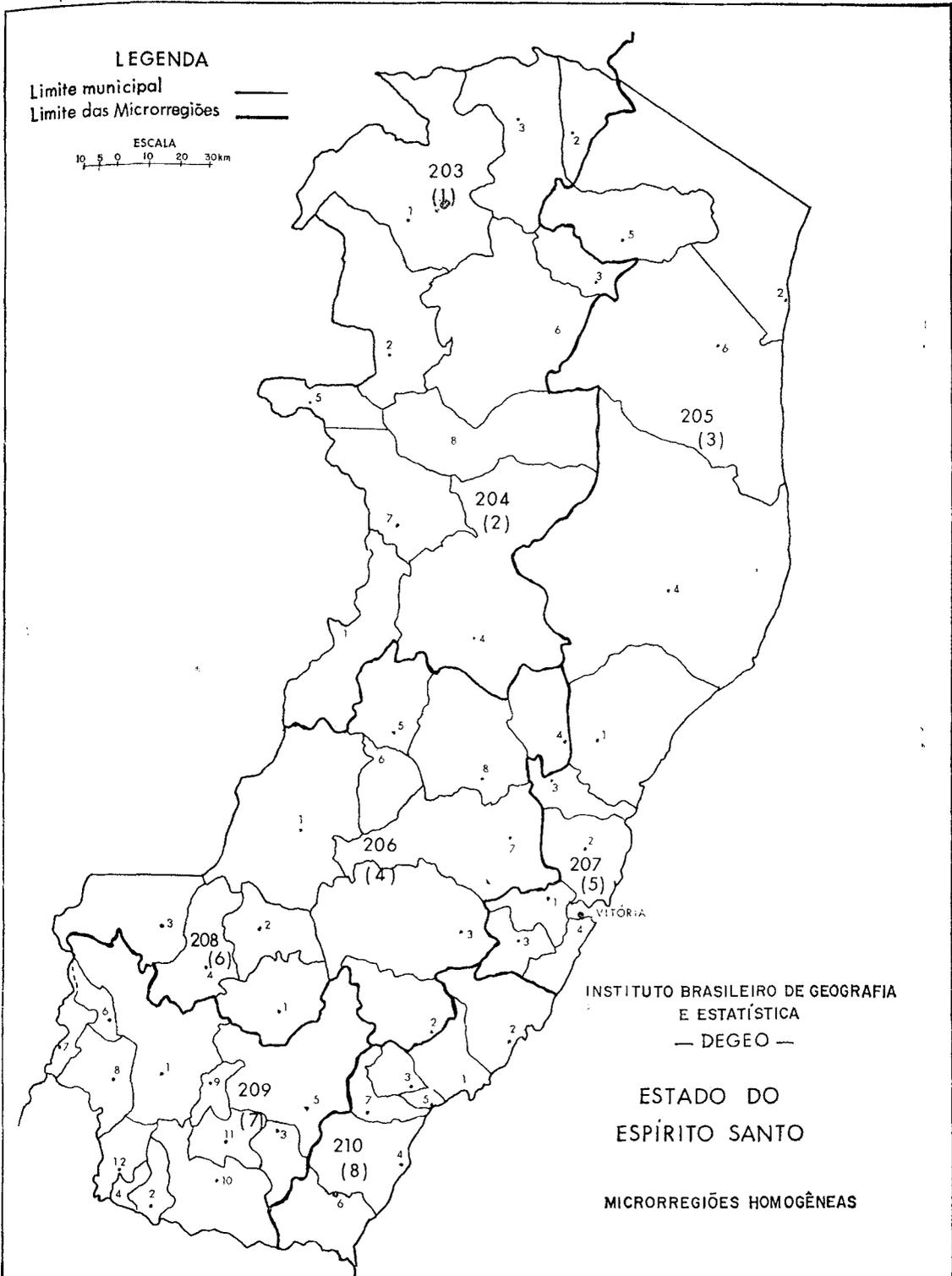


FIGURA 7

Mapa do Estado do Espírito Santo com a delimitação das microrregiões e respectivos municípios



desta variável uma vez que a sua posição permite aparentemente uma explicação sensata. Dir-se-ia que o dado está inconsistente. Recorrendo-se à Figura 2, que nos dá a “qualidade da representação”, lembramos que a contribuição de 003 a inércia é a mais elevada na nuvem de colunas. Isso em si não é indicativo de erro, simplesmente, como se disse, o é de que este ponto possui um comportamento atípico na tipologia gerada pelos municípios. No entanto, examinando-se a tabela de contingência na Fig. 3, encontra-se, como era de se esperar pelo gráfico que esta variável está assumindo valores altos nas micro 5 e 8 com relação aos que assume nas restantes. Identifica-se com alta probabilidade um erro no levantamento dessa variável uma vez que a realidade é justamente inversa (vide [6], [8]). Tem-se aqui um exemplo da utilidade da análise de correspondências como instrumento de crítica qualitativa de um conjunto de dados.

Façamos uma rápida análise do plano fatorial 1×3 . Os pontos 005 e 804 caracterizam-se mais nitidamente ainda como valores extremos (out liers). A menor variância do eixo 3 ($= 0,134869$) efetua uma compressão horizontal dos pontos acentuando-se então a separação evidente, pela reta de cota 0.000715, entre as regiões 5 e 8 e o restante. A faixa entre ela e sua paralela de cota 0.000059 é uma zona de transição contendo alguns municípios da 7ª região, para depois aparecer o aglomerado agrícola. Mais evidente também fica o estreito relacionamento das variáveis 001, 002, 004, 006, 007 com o grupo inferior e das de código 008, 009, 010, 011, 012, 013 com o grupo superior.

QUADRO 2

VALOR DO RAI DO CÍRCULO DE GUARDA PARA ALGUNS PONTOS

(A fórmula do raio do círculo de guarda do ponto j é, como visto na

Parte I, $\sqrt{\frac{6}{n_j}}$)

	raio	dist. origem
município 102:	$(6 \times 9356^{-1})^{1/2} \cong 10^{-2} \times 2,5$	0.002888
município 103:	$(6 \times 5464^{-1})^{1/2} \cong 10^{-2} \times 3,3$	0.002409
município 804:	$(6 \times 2189^{-1})^{1/2} \cong 10^{-2} \times 5,2$	0.003492
variável 3:	$(6 \times 7214^{-1})^{1/2} = 0.028839$	0.003942
variável 5:	$(6 \times 2999^{-1})^{1/2} = 0.044729$	0.003089
variável 7:	$(6 \times 47343^{-1})^{1/2} = 0.01$	0.001438

2.5 Procuramos assim ilustrar os passos e a mecânica da interpretação de um resultado. De forma alguma esgotou-se as interpretações possíveis de se extrair do exemplo proposto nem se pretende que as condições extraídas e a maneira como se a elas chegou sejam aceitas como conceitos rígidos. O objetivo é transmitir o sentimento de como se deve utilizar o método. Os princípios gerais, como acontece de hábito, apareceram: o primeiro eixo revelou as oposições básicas entre os dados (sejam linhas ou colunas) ao passo que o segundo revelou diferenças dentro de dois grandes grupos, através o princípio do baricentro o inter-relacionamento das nuvens foi realizado, e assim por diante. Não se pense também que o trabalho pararia aqui, uma nova análise deve ser feita retirando-se as variáveis 003 e 005, o município 804 bem como o 505 (Vitória) a fim de se obter, provavelmente, uma distribuição do novo universo.

BIBLIOGRAFIA

- 1 — ANDERSON, T. W. *An Introduction to Multivariate Statistical Analysis*. New York, John Wiley, 1958. 374 p.
- 2 — BENZÉCRI, J-P et alii. *L'Analyse de Données*. Paris, Dunod, 1973. 2v.
- 3 — BURTSCHY, Bernard. *Analyse des Données: rappell théorique, programmes d'utilisation*. Paris, Direction Generale des Telecommunications Groupement des Études Economiques et Informatiques, 1973. 83 p.
- 4 — DRUMMOND JR., M. E. *Evaluation and Measurement Techniques for Digital Computer Systems*. Englewood Cliffs, N. J., Prentice-Hall, 1973. 338 p.
- 5 — FUNDAÇÃO IBGE, Rio de Janeiro. Departamento de Censos. *Censo Agropecuário — Espírito Santo*. Rio de Janeiro, 1974 (VIII Recenseamento Geral — 1970 Série Regional v. 3, t. 15).
- 6 — ———. *Censo Demográfico — Espírito Santo*, Rio de Janeiro, 1973 (VIII Recenseamento Geral — 1970. Série Regional v. 1, t 15).
- 7 — ———. *Censo Demográfico — Guanabara*. Rio de Janeiro, 1973 (VIII Recenseamento Geral — 1970. Série Regional v. 1, t. 17).
- 8 — ———. *Censo Industrial — Espírito Santo*. Rio de Janeiro, 1973 (VIII Recenseamento Geral — 1970. Série Regional v. 4, t. 15).
- 9 — GOODMAN, Leo. A General Model for the Analysis of Surveys. *The American Journal of Sociology*, 77 (6):1035-86, May 1972.
- 10 — ———. *Everymans's Contingency Tables Analysis. A. Statistical Package Available at the Departament of Statistics*. Chicago, University of Chicago.
- 11 — HAWKINS, Douglas M The Detection of Errors in Multivariate Data Using Principal Components. *JASA*, 69 (346):340-4, June 1974.
- 12 — INSTITUTO BRASILEIRO DE GEOGRAFIA, Rio de Janeiro *Divisão do Brasil em Microrregiões Homogêneas*. Rio de Janeiro, 1970
- 13 — KRISHNAIAH, P. R., ed. *Multivariate Analysis*. New York, Academic Press, 1966. 592 p
- 14 — ———. *Multivariate Analysis II*. New York, Academic Press, 1969. 696 p.
- 15 — LEBART, L. et alii. *Recherches sur la Description Automatique des Données Socio-Economiques*. Paris, Centre de Recherches et Documentation sur la Consommation, 1969.

- 16 — NOBLE, Ben. *Applied Linear Algebra*. Englewood Cliffs N. J., Prentice-Hall, 1969.
- 17 — PEARSON, K. On a Criterion that a Given System of Derivations from the Probable in the Case of a Correlated System of Variables is such that it Can Reasonably Be Supposed to Have Arisen in Random Sampling. *Phil. Mag.* (5):50-157.
- 18 — ——— On Lines and Planes of Closest Fit to Systems of Points in Space. *Phil. Mag.* 3 (2):559-72, June 1901.
- 19 — RAO, C. R. The Use and Interpretation of Principal Component Analysis in Applied Research. *Sankhyā*.
- 20 — SADE. Sistema de Análise de Dados e Estatística. Versão 1. Rio de Janeiro, IBGE. Centro de Informática, 1974.
- 21 — SHARPE, W. F. *The Economics of Computers*. New York, Columbia University Press, 1969.
- 22 — TUKEY, J. W. The Future of Data Analysis. *Annals of Mathematical Statistics*.
- 23 — ———. & WILK, M. B. *Data Analysis and Statistics: Techniques and Approaches*. California, 1965. Proceedings of the Symposium on Information Processin in Light Sensory Systems, Nov. 1-3, 1965.
- 24 — VALLE, Nelson do. *Posição social das ocupações*. Rio de Janeiro, IBGE. Centro de Informática, 1974.

TENDÊNCIA DA MORTALIDADE POR CAUSAS NO BRASIL*

João Yunes

da Faculdade de Saúde Pública da USP

Jorge Somensi

da Faculdade de Saúde Pública da USP

e

Vera Shirley C. Ronchezel

da Secretaria de Saúde do Estado de S. Paulo

SUMÁRIO

Resumo

1. *Introdução*
2. *Objetivos*
3. *Material e Método*
4. *Resultado e discussão*
 - 4.1 *Região Norte*
 - 4.2 *Região Nordeste*
 - 4.3 *Região Sudeste*
 - 4.4 *Região Centro-Oeste*
 - 4.5 *Região Sul*
 - 4.6 *Brasil*
5. *Conclusões*

Referências Bibliográficas

Anejos: Gráficos e Tabelas

* Trabalho apresentado pelos autores no Encontro Brasileiro de Estudos Populacionais, de 29 de julho a 3 de agosto de 1974, promovido pelo IBGE, na cidade do Rio de Janeiro.

RESUMO

Através do estudo da mortalidade por causas mal definidas, doenças transmissíveis, tuberculose e neoplasias, segundo alguns indicadores, procurou-se estabelecer a evolução do nível de saúde nas regiões fisiográficas brasileiras e no País como um todo, no período de 1959 a 1970. Para cada grupo dessas doenças estabeleceu-se seu coeficiente de mortalidade por 100 mil habitantes, a proporção de óbitos no conjunto de todas as causas, e, para a tuberculose, sua importância proporcional dentro do grupo das doenças transmissíveis.

Concluiu-se serem ainda, muito importantes na mortalidade as doenças transmissíveis e as causas mal definidas.

A análise dos indicadores faz crer ter havido deterioração do nível de saúde no Brasil nos últimos anos.

1. INTRODUÇÃO

Algumas causas de óbito quando analisadas numa série histórica e submetidas a um enfoque adequado permitem avaliar a evolução das condições de saúde de uma determinada área. Assim é que nos propuemos a estudar como se comportou a mortalidade por doenças transmissíveis (e destas especificamente por tuberculose), por neoplasias, e por causas mal definidas nas Regiões Fisiográficas do Brasil, numa série histórica de onze anos, com a finalidade de verificar, através destes indicadores, as condições de saúde nas diversas regiões fisiográficas e no País como um todo.

Julgamos que as quatro causas selecionadas atendem ao propósito final do trabalho. As doenças transmissíveis certamente indicarão melhora do estado sanitário se sua força na mortalidade geral for decrescendo com o tempo, já que, de um modo geral, são doenças controláveis através do saneamento adequado do meio, das imunizações e da educação sanitária. As doenças degenerativas, e dentre elas as neoplasias, tenderão a substituir as doenças transmissíveis como causa de morte, conforme a melhora das condições sanitárias da área for se verificando. Da mesma forma, a mortalidade por causas mal definidas nos dá indicação da qualidade e da cobertura dos recursos de saúde. Das doenças transmissíveis deu-se ênfase à mortalidade por tuberculose, apesar de sabermos ser mais conveniente o estudo de sua morbidade (incidência e prevalência). No entanto, a não disponibilidade de dados de morbidade fez-nos basear nosso estudo na mortalidade, o que, de certa forma, nos dá indicação da primeira.

É evidente que para se ter melhor idéia da evolução das condições sanitárias do País, outros indicadores poderiam ser analisados, mas a limitação das informações e a qualidade dos dados dificultam seu uso. Por essas razões, não se pretende que as afirmações aqui expostas reflitam a verdade absoluta, mas sim dêem uma visão panorâmica da problemática de saúde no Brasil.

2. OBJETIVOS

2.1 Estudar a evolução da mortalidade por doenças transmissíveis, tuberculose, neoplasias e causas mal definidas nas Regiões Fisiográficas do Brasil e no País como um todo.

2.2 Com base em alguns indicadores de mortalidade por essas doenças, verificar como se comportaram as condições de saúde no Brasil e em suas Regiões Fisiográficas.

3. MATERIAL E MÉTODO

Os dados foram obtidos do “Anuário Estatístico do Brasil”, (2-13) e “Séries Estatísticas Retrospectivas — 1970”, todas publicações do Instituto Brasileiro de Geografia e Estatística. Os dados referem-se às capitais dos Estados pois eles não são disponíveis para o Estado como um todo. Portanto, quando apresentamos dados para as Regiões Fisiográficas e Brasil eles se referem a informações das capitais. Os registros de óbitos aparecem nas publicações do IBGE segundo os grandes grupos de causa (Lista B), segundo o número de óbitos ocorrido e o coeficiente por 100 mil habitantes, para cada capital estadual. Os dados apresentados por Região Fisiográfica foram obtidos da soma dos óbitos ocorridos nas capitais componentes da respectiva região fisiográfica e relacionados à sua população.

As capitais de Estados e Territórios que compõem cada Região Fisiográfica são as seguintes:

- a) Região Norte: Rio Branco, Porto Velho, Manaus, Boa Vista, Belém e Macapá. Só há registros para Belém e Manaus;
- b) Região Nordeste: São Luís, Teresina, Fortaleza, Natal, João Pessoa, Recife, Maceió, Aracaju e Salvador;
- c) Região Centro-Oeste: Cuiabá, Goiânia e Brasília;
- d) Região Sudeste: Niterói, Belo Horizonte, Vitória, São Paulo, Rio de Janeiro;
- e) Região Sul: Curitiba, Florianópolis e Porto Alegre.

Para cada doença estudada foi determinado o coeficiente específico de mortalidade referido a 100 mil habitantes. Foi, também, estabelecida a proporção de óbitos que cada doença enfocada representa sobre os óbitos por todas as causas. Para a tuberculose foi determinado, ainda, o percentual de óbitos por essa causa sobre os óbitos por todas as doenças transmissíveis, ficando demonstrada, assim, sua força dentro da mortalidade por enfermidades infecciosas.

Em alguns anos, como não se dispunha de dados para determinada capital, os índices e coeficientes acima expostos foram calculados tomando-se como referência a população e os óbitos das demais capitais componentes da região. Registramos que não há dados de mortalidade para Belém em 1968; para Fortaleza em 1964; para o Rio de Janeiro em 1959, 1963 a 1967 e 1970; para São Paulo de 1968 a 1970; para Cuiabá em 1960 e 1969; para Brasília de 1959 a 1962; para Curitiba em 1967; para Florianópolis em 1959, 1961, 1962, 1966 e 1968 a 1970; para Porto Alegre em 1959 e 1964 a 1966.

No presente trabalho utilizou-se a Classificação Internacional de Doenças (8.^a Revisão — 1965). Desse modo, “doenças transmissíveis” inclui todas as enfermidades infecciosas (de B-1-B-18, segundo a classificação pela lista B); “tuberculose” inclui a do aparelho respiratório e outras formas, inclusive os efeitos tardios (B-5 a B-6 da lista B); “neoplasias” inclui os tumores malignos dos tecidos linfáticos e dos órgãos hematopoiéticos, tumores benignos e de natureza não especificada (B-19 a B-20 da lista B); “causas mal definidas” inclui sintomas e estados mórbidos mal definidos (B-45).

A análise da mortalidade pelas causas citadas foi feita numa série histórica de onze anos, de 1959 a 1970, uma vez que para anos anteriores a 1959 não há registros oficiais de mortalidade para a maioria das capitais.

4. RESULTADO E DISCUSSÃO

4.1 Região Norte

O Gráfico 1 representa a variação dos coeficientes específicos de mortalidade por doenças transmissíveis, tuberculose, neoplasias e causas mal definidas. O coeficiente da mortalidade por doenças transmissíveis era, em 1959, pouco maior do que 200/100 mil habitantes, foi decrescendo até atingir o menor nível em 1967, quando chegou a 116,4/100 mil habitantes, ocorrendo, em 1968, importante aumento, chegando a 331,8/100 mil habitantes, decrescendo a seguir, em 1969, a nível semelhante ao alcançado em 1959, estando em torno de 200/100 mil habitantes em 1970. Vê-se que esse indicador sofreu, em níveis sempre altos, alterações bruscas no decorrer do período, que podem ser devidas a várias causas: a) alterações próprias na mortalidade por doenças transmissíveis; b) flutuações no rigor com que esses dados foram registrados e analisados; c) o pico ocorrido em 1968 pode ser devido, em parte, a não inclusão de dados de Belém, ficando, portanto, a referência só a Manaus, cuja mortalidade por doenças transmissíveis é bem mais elevada do que a de Belém. Observa-se, de acordo com o Gráfico 1, que este coeficiente, apesar de ser elevado, praticamente não sofreu alterações depois de decorridos 11 anos. Na Tabela 1 mostra-se a mortalidade proporcional por doenças transmissíveis na Região Norte e nas demais Regiões Fisiográficas do Brasil.

O coeficiente de mortalidade por tuberculose (Tabela 2 e Gráfico 1) estava acima de 100/100 mil habitantes em 1959 e decresceu para pouco menos da metade em 1970. Até 1965 este decréscimo foi mais ou menos paralelo ao das doenças transmissíveis, após esse ano houve uma dissociação entre as duas curvas, mostrando não só que a parcela da tuberculose dentro das doenças transmissíveis foi diminuindo, mas também que a mortalidade pelas outras doenças transmissíveis tornou-se mais importante.

A comparação dos coeficientes de mortalidade por tuberculose e por neoplasias (Tabela 3 e Gráfico 1) nos dá uma idéia do precário nível de saúde desta região, pois uma doença transmissível, a tuberculose, com métodos diagnósticos, terapêuticos e profiláticos conhecidos e eficazes, contribui para o obituário geral de forma tão intensa quanto a de todas as neoplasias, doenças de poucos recursos preventivos e terapêuticos, na maioria dos casos.

A observação do Gráfico 1 nos mostra que a diferença entre os dois coeficientes era de cerca de 50/100 mil habitantes em 1959, paulatinamente as duas curvas foram convergindo até dar-se o cruzamento por volta de 1968, sendo que em 1970 a diferença entre os coeficientes era de aproximadamente 16/100 mil habitantes em favor das neoplasias. No entanto essa superação da mortalidade por neoplasias em relação à tuberculose não indica, por si só, melhoria das condições de saúde da área, pelo contrário, elas parecem bastante deterioradas tendo em vista o alto coeficiente de mortalidade por todas as doenças transmissíveis.

Quanto aos óbitos por causas mal definidas (Gráfico 1 e Tabela 4), seu coeficiente é baixo em relação aos outros três coeficientes analisados, mas a tendência é de crescimento, estando no último ano do período quase no mesmo nível de há 11 anos atrás.

4.2 Região Nordeste

Os mesmos coeficientes específicos e índices estão representados nas Tabelas 1 a 5 e no Gráfico 2 para a Região Nordeste.

O coeficiente de mortalidade por doenças transmissíveis (Gráfico 2), que em 1959 era bastante elevado (cerca de 270/100 mil habitantes), sofreu importante e sistemático decréscimo até 1966/67, elevando-se a seguir e bruscamente a partir de 1968, estando estabilizado nos dois últimos anos do período nos mesmos níveis de 1959. A mortalidade proporcional por esse grupo de doenças, que de 1959 a 1968 oscilou entre 11 e 16%, em 1969/70 representou cerca de 24% de todos os óbitos (Tabela 1).

A mortalidade por tuberculose em 1959 apresentou um coeficiente de quase 100 óbitos por 100 mil habitantes, havendo a partir desse ano um progressivo decréscimo até cerca de 50/100 mil habitantes em 1970. Observa-se aqui também uma diminuição na parcela da tuberculose dentre as doenças transmissíveis.

Comparando este coeficiente com o de mortalidade por neoplasias, vê-se a importante diferença entre os dois no início do período (cerca de 50/100 mil habitantes em favor da tuberculose) e em 1970 a diferença é de aproximadamente 20/100 mil habitantes em favor das neoplasias. O cruzamento das duas curvas deu-se em 1966 em consequência do decréscimo de um dos coeficientes (tuberculose) e aumento, menos importante, do outro (neoplasias).

Chama bastante atenção o comportamento histórico, nesta região, do coeficiente de mortalidade por causas mal definidas (Gráfico 2 e Tabela 4). Seu valor era extremamente elevado em 1959 (cerca de 250/100 mil habitantes), decresceu rapidamente até 1962, quando chegou a 116/100 mil habitantes, sofreu oscilações em 1963/64, subindo bruscamente para mais de 200/100 mil habitantes em 1965 e a partir daí decresceu levemente, estando em 184/100 mil habitantes em 1970.

Provavelmente o valor baixo deste coeficiente em 1964 deve-se a não inclusão de dados, para esse ano, de Fortaleza, cidade que apresentou cerca de 50% dos óbitos por causas mal definidas nos anos em torno de 1964.

Os valores elevados dos coeficientes de mortalidade por doenças transmissíveis e por causas mal definidas e o baixo coeficiente de mortalidade por neoplasias nos indicam que a situação de saúde desta região é precária e que não houve melhora no decorrer dos anos analisados.

4.3 Região Sudeste

Contrariamente às duas regiões já analisadas, nesta não sobressaem tão nitidamente as doenças transmissíveis como causa de morte (Tabela 1 e Gráfico 3). De fato, seu coeficiente de mortalidade é aproximadamente igual ao das neoplasias no período de 1959/62, sendo este superior ao primeiro, nitidamente, de 1962/68, ano em que seus valores são coincidentes. A partir daí as doenças transmissíveis cresceram acentuadamente, sendo em 1970 a diferença entre os dois coeficientes de mais de 50/100 mil habitantes. Talvez, para esta região, a situação desses dois coeficientes ficasse alterada caso fossem, para todos os anos da série, incluídos dados do Rio de Janeiro e de São Paulo, as duas capitais mais importantes da região, já que seria de esperar um maior número de óbitos por neoplasias, pelos recursos médicos que possuem, ocasionando convergência para elas de óbitos por câncer.

O coeficiente de mortalidade por tuberculose diminuiu de 1959 a 1967, elevando-se discretamente a partir desse ano, estando atualmente em torno de 40/100 mil habitantes (Gráfico 3).

Nesta região, na série histórica considerada, percebe-se uma quase superposição das curvas representativas dos coeficientes de mortalidade por doenças transmissíveis e por neoplasias e, em valores menores, uma superposição das curvas que representam os coeficientes de mortalidade por causas mal definidas e por tuberculose. Nota-se,

ainda, um certo paralelismo entre as curvas dos coeficientes de mortalidade por doenças transmissíveis e por tuberculose, de 1959 a 1969, indicando que a forma da tuberculose dentro das doenças transmissíveis acompanhou a mortalidade por todas essas doenças. A partir de 1969 houve dissociação das duas curvas, principalmente pelo aumento do coeficiente de mortalidade por doenças transmissíveis.

4.4 Região Centro-Oeste

Na observação do Gráfico 4 chama atenção o comportamento do coeficiente de mortalidade por causas mal definidas. Este coeficiente era extremamente elevado no início do período (cerca de 240/100 mil habitantes em 1959), houve uma redução importante de 1961 a 1963, estabilização em valores próximos a 130/100 mil habitantes de 1963/67 e em 1967/68 uma grande redução, estando, em 1970, próximo de 35/100 mil habitantes. Portanto, em 11 anos, o coeficiente de mortalidade por causas mal definidas sofreu um decréscimo de 170/100 mil habitantes, sugerindo melhora de cobertura da assistência médica.

O coeficiente de mortalidade por doenças transmissíveis sofreu bruscas oscilações, explicáveis pela deficiência das estatísticas vitais. A partir de 1967, a mortalidade por doenças transmissíveis apresenta tendência a elevar-se, tendo, em 1970, coeficiente próximo a 200/100 mil habitantes, valor maior do que no início do período em análise.

O coeficiente de mortalidade por neoplasias, em todo o período, foi maior do que o da mortalidade por tuberculose. A partir de 1964 começa a haver lenta dissociação entre as duas curvas, principalmente pela queda do valor do coeficiente de mortalidade por tuberculose (Gráfico 4).

Nesta região a importância da mortalidade por tuberculose, relativamente às doenças transmissíveis, diminuiu significativamente com o passar dos anos, especialmente pelo aumento da mortalidade pelas outras doenças transmissíveis (Tabela 2).

4.5 Região Sul

De 1959 a 1965, praticamente, os coeficientes de mortalidade por doenças transmissíveis e por neoplasias tiveram o mesmo valor (Gráfico 5). A partir de 1965 as neoplasias foram adquirindo maior importância, sendo para esta região a principal causa de óbito dentre as quatro consideradas.

O coeficiente de mortalidade por tuberculose foi bastante elevado de 1959 a 1963, diminuindo a partir de 1964 e estabilizando-se em seguida a este ano, apesar de ser ainda alto (31,9/100 mil habitantes em 1970). Deve ser ressaltada a participação proporcional da tuberculose na mortalidade por doenças transmissíveis. Assim, nos anos iniciais da década de 60, mais de 60% dos óbitos por doenças transmissíveis foram por tuberculose. Esta porcentagem foi decrescendo gradativamente

até ficar em 41,5% (Tabela 2). Este fato nos dá indicação de que, além da elevada mortalidade por tuberculose, a mortalidade por outras doenças transmissíveis não é tão importante, sugerindo melhor saneamento do meio, maior cobertura de imunização contra doenças transmissíveis e nível sócio-econômico mais elevado do que em outras regiões.

A mortalidade por causas mal definidas apresenta um coeficiente que se elevou bruscamente a partir de 1962, atingindo o valor máximo (mais de 150/100 mil habitantes) em 1964. A partir deste ano houve um decréscimo acentuado, estando em 26,4/100 mil habitantes em 1970, que é o menor dos coeficientes apresentados na série histórica estudada.

Deve ser ressaltada a precariedade da constância dos dados estatísticos de mortalidade na Região Sul. Isto é tão evidente que para 1967 nenhuma das três capitais componentes da Região apresentava registros oficiais de óbitos.

4.6 Brasil

O Gráfico 6 e a Tabela 5 referem-se à mortalidade por doenças transmissíveis, por tuberculose, por neoplasias e por causas mal definidas no Brasil como um todo, a partir de dados das capitais estaduais e da Capital Federal.

Novamente aqui sobressaem as doenças transmissíveis com um coeficiente superior a 150/100 mil habitantes em 1959, mas que diminui paulatinamente até 1967 voltando a elevar-se a partir deste ano. Seu coeficiente em 1970 foi maior do que 200/100 habitantes, valor bem superior ao registrado no ano inicial da série em estudo.

A curva representativa do coeficiente de mortalidade por tuberculose praticamente acompanhou a tendência de queda das doenças transmissíveis até 1967, quando se verificou uma dissociação da tendência pela ascensão da curva das doenças transmissíveis e queda dos óbitos por tuberculose.

A diminuição da mortalidade por tuberculose não indica, necessariamente, um equacionamento adequado do problema, uma vez que os recursos terapêuticos atuais fazem esperar uma diminuição da mortalidade por essa causa. Dessa forma, seria mais elucidativo conhecer a incidência e a prevalência da doença.

Almeida, A. P. de et al (16) conclui que a tuberculose no Brasil nos últimos 10 anos (1960/1969) tem apresentado menores coeficientes de mortalidade em quase todas as capitais e territórios e possivelmente também no interior; e não possuímos ainda indicadores epidemiológicos aceitáveis quanto à incidência e prevalência dos casos ou índices de infecção que possam ser aplicáveis a todo o Brasil; e os dados relativos à inscrição de casos novos e doentes existentes nos dispensários quer nas capitais quer no interior do País, mostram que apesar dos recursos da "quimioterapia standard", o número de doentes matriculados nessas

unidades não tem sofrido modificação apreciável, a não ser nos últimos 3 anos (1967-1969) em que as taxas de inscritos no ano e total de doentes matriculados foram um pouco mais baixas.

O coeficiente de mortalidade por causas mal definidas apresenta-se acima de 100 óbitos por 100 mil habitantes em 1959, diminui até cerca de 50/100 mil habitantes em 1962 e a partir daí cresce constantemente, situando-se acima de 100/100 mil habitantes em 1970.

A mortalidade por neoplasias mostra-se praticamente estável durante o período, com um coeficiente em torno de 80/100 mil habitantes. Em 1970 seu valor apresenta-se menor do que os de mortalidade por causas mal definidas e por doenças transmissíveis, sendo que para este último a diferença é superior a 130%, isto é, o risco de morrer por doenças transmissíveis, em 1970, no Brasil, era duas vezes superior ao risco de morrer por câncer.

Ao se analisar as enfermidades por região, observa-se que na Nordeste, conforme Gráfico 7, os coeficientes de mortalidade por doenças transmissíveis estiveram sempre maiores que nas demais regiões, exceto nos anos de 1964, 1966 e 1968 em que se apresentaram pouco menores do que na Região Norte. Os menores coeficientes correspondem à Região Sudeste, exceto para os três últimos anos estudados em que se apresentam maiores que para a Sul (Tabela 1).

A mortalidade por tuberculose, de acordo com o Gráfico 8, apresenta-se maior em todos os anos na Região Norte, seguida pela Região Nordeste. Os menores coeficientes são encontrados na Região Centro-Oeste na série histórica estudada.

A contribuição da tuberculose no coeficiente de mortalidade por doenças transmissíveis foi sempre maior na Região Norte do que na Região Nordeste em todos os anos estudados, tanto que em 1970, na Norte, 23,8% dos óbitos por doenças transmissíveis eram devidos à tuberculose, contra apenas 16,3% para a Nordeste. Entretanto, a tuberculose como parcela de causa de óbitos por doenças transmissíveis é muito mais importante na Região Sul que em qualquer outra (em 1970, 41,6% dos óbitos por doenças transmissíveis são devidos à tuberculose). A região onde a tuberculose apresenta menor proporção como doença transmissível é a Região Centro-Oeste (em 1970, 7,3).

A mortalidade por causas mal definidas é maior na Região Nordeste que em qualquer outra. Enquanto que nesta o coeficiente se apresenta igual a 184,6/100 mil habitantes em 1970, nas demais regiões, no mesmo ano, está em torno de 50/100 mil habitantes. A situação é mais ou menos a mesma ao longo de toda a série histórica, exceto nos quatro primeiros anos em que a Região Centro-Oeste apresenta coeficientes maiores, e em 1964 a Região Nordeste tem valores menores que as das Regiões Sul e Centro-Oeste. Nos demais anos a situação é mais ou menos a mesma observada para 1970 (Gráfico 9).

A mortalidade por neoplasias malignas esteve sempre maior para as Regiões Sul e Sudeste que para as demais. Assim é que enquanto as

duas primeiras apresentavam coeficientes em torno de 100 óbitos por 100 mil habitantes nas demais os valores estiveram em torno de 50/100 mil habitantes em todos os anos considerados (Gráfico 10). Observa-se, ainda, que os coeficientes foram maiores para a Região Sul que para a Sudeste (exceto nos anos de 1959 e 1968) e que a partir de 1963 os menores coeficientes apresentados correspondem à Região Centro-Oeste, seguida da Região Norte que apresentava os menores coeficientes até 1962.

5. CONCLUSÕES

1. Das causas de óbito analisadas, as determinadas por doenças transmissíveis apresentam os maiores coeficientes de mortalidade para todas as regiões, exceto na Região Sul em que predominam as neoplasias.
2. Constata-se um nítido aumento da mortalidade por doenças transmissíveis a partir de 1967, em todas as regiões, exceto na Sul.
3. As causas do aumento da mortalidade por doenças transmissíveis podem sugerir:
 - a) piora do nível sócio-econômico da população;
 - b) migração intensa de populações rurais para zonas urbanas;
 - c) difícil acesso de grande parte da população a recursos de saúde;
 - d) deterioração dos serviços de saúde;
 - e) melhoria da informação da causa básica da morte.
4. Há um desnível, quanto à situação de saúde, entre as diversas regiões fisiográficas brasileiras. A que apresenta nível mais satisfatório é a Região Sul e a de pior nível é a Nordeste.
5. A análise dos indicadores utilizados faz concluir que houve deterioração do nível de saúde no Brasil nos últimos anos.

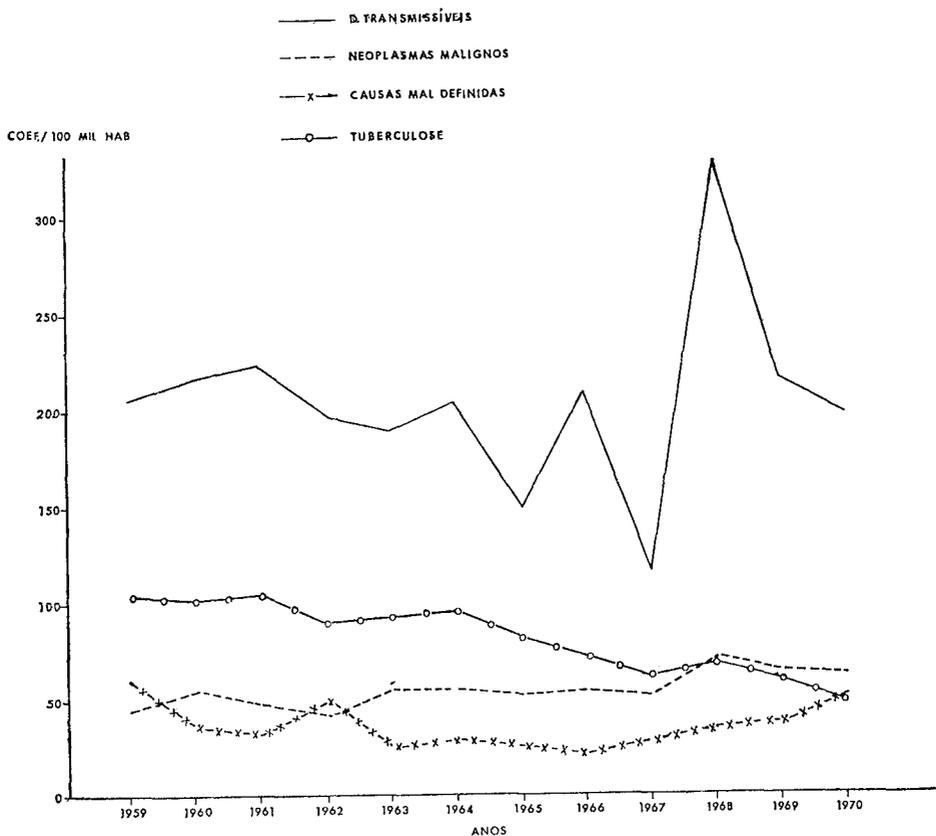
REFERÊNCIAS BIBLIOGRÁFICAS

- 1 — ALBUQUERQUE, A. F. R. de & VILLAS BOAS, A. Informações epidemiológicas sobre a tuberculose nas capitais brasileiras. Separata da *Revista do Serviço Nacional de Tuberculose*, 10 (37) 1966
- 2 — ANUÁRIO ESTATÍSTICO DO BRASIL. Rio de Janeiro, Fundação IBGE, v. 23-33, 1961-1972
- 3 — CAMPOS, O. A tuberculose no Brasil — evolução e tendência. *Revista do Serviço Nacional de Tuberculose*, 11:229-35
- 4 — FUNDAÇÃO IBGE, Rio de Janeiro. *Brasil: Séries estatísticas retrospectivas*. Rio de Janeiro, 1970. v. 1
- 5 — ALMEIDA, A. P. de & LIMA, L. L. de. A tuberculose no Brasil (1950-1970) *Revista do Serviço Nacional de Tuberculose*, 17 (66):141-54, 1973

ANEXO: GRÁFICOS E TABELAS

GRÁFICO 1

MORTALIDADE POR DOENÇAS TRANSMISSÍVEIS, POR TUBERCULOSE, POR NEOPLASIAS E POR CAUSAS MAL DEFINIDAS (COEFICIENTE POR 100 MIL HABITANTES) — REGIÃO NORTE — 1959 A 1970

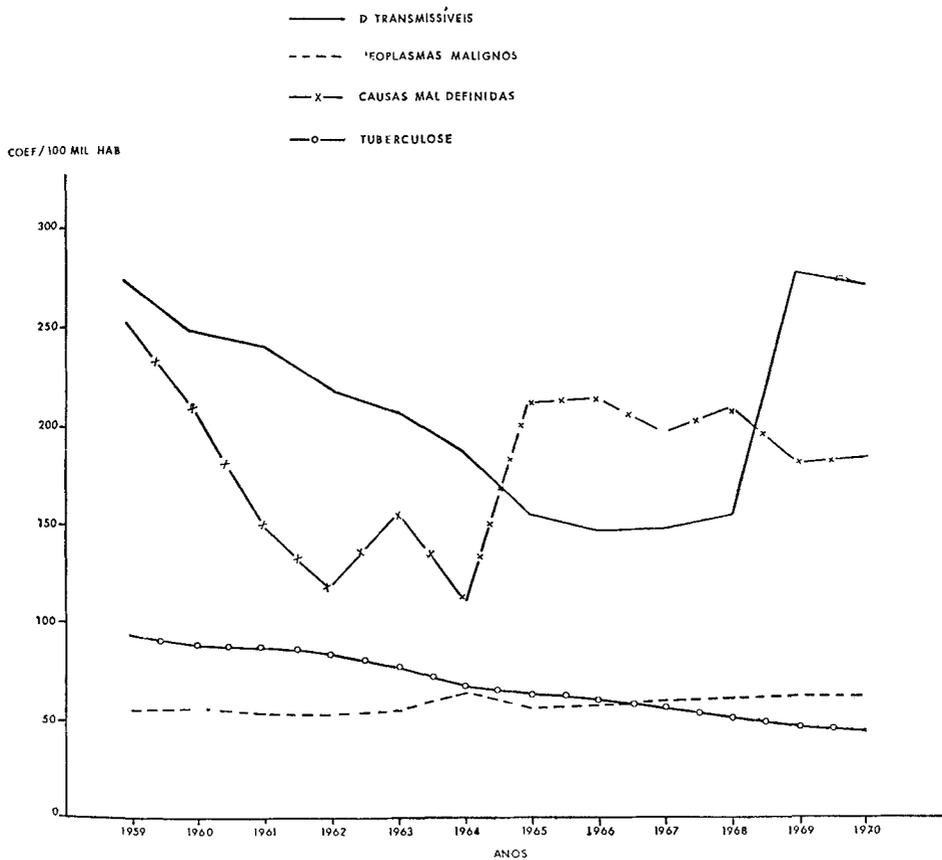


FONTE: 1 — Anuário Estatístico do Brasil (2 — 12)

2 — Séries Estatísticas Retrospectivas — 1970 (14)

GRÁFICO 2

MORTALIDADE POR DOENÇAS TRANSMISSÍVEIS, POR TUBERCULOSE, POR NEOPLASIAS E POR CAUSAS MAL DEFINIDAS (COEFICIENTE POR 100 MIL HABITANTES) — REGIÃO NORDESTE — 1959 A 1970

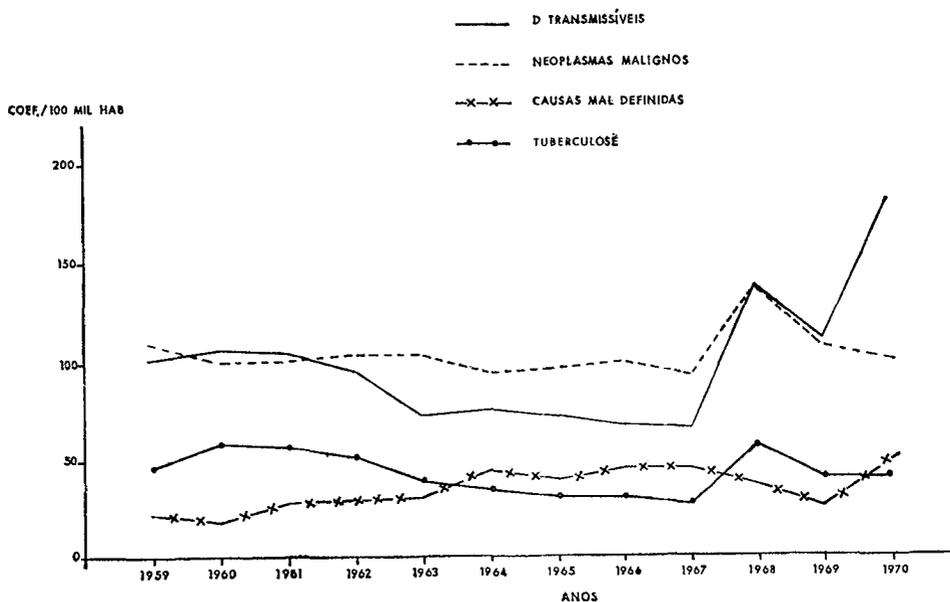


FONTE: 1 — Anuário Estatístico do Brasil (2 — 12)

2 — Séries Estatísticas Retrospectivas — 1970 (14)

GRÁFICO 3

MORTALIDADE POR DOENÇAS TRANSMISSÍVEIS, POR TUBERCULOSE, POR NEOPLASIAS E POR CAUSAS MAL DEFINIDAS (COEFICIENTE POR 100 MIL HABITANTES) — REGIÃO SUDESTE — 1959 A 1970

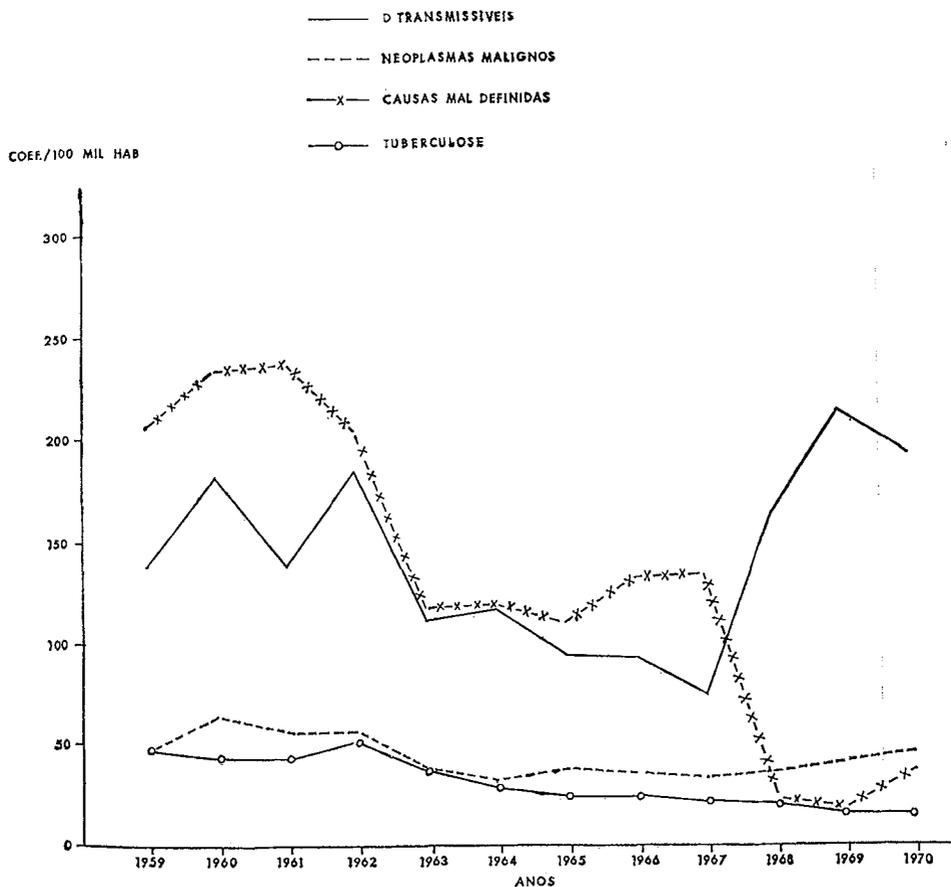


FONTE: 1 — Anuário Estatístico do Brasil (2 — 12)

2 — Séries Estatísticas Retrospectivas — 1970 (14)

GRÁFICO 4

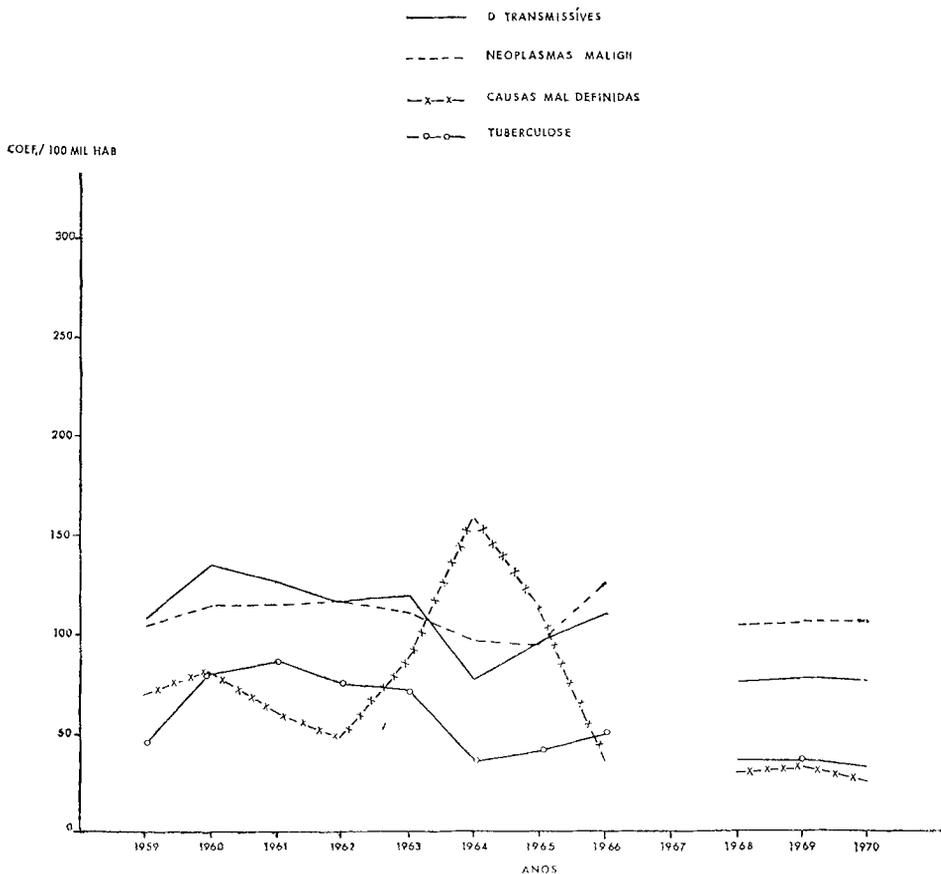
MORTALIDADE POR DOENÇAS TRANSMISSÍVEIS, POR TUBERCULOSE, POR NEOPLASIAS E POR CAUSAS MAL DEFINIDAS (COEFICIENTE POR 100 MIL HABITANTES) — REGIÃO CENTRO-OESTE — 1959 A 1970



FONTE: 1 — Anuário Estatístico do Brasil (2 — 12)
2 — Séries Estatísticas Retrospectivas — 1970 (14)

GRAFICO 5

MORTALIDADE POR DOENÇAS TRANSMISSÍVEIS, POR TUBERCULOSE, POR NEOPLASIAS E POR CAUSAS MAL DEFINIDAS (COEFICIENTE POR 100 MIL HABITANTES) — REGIÃO SUL — 1959 A 1970

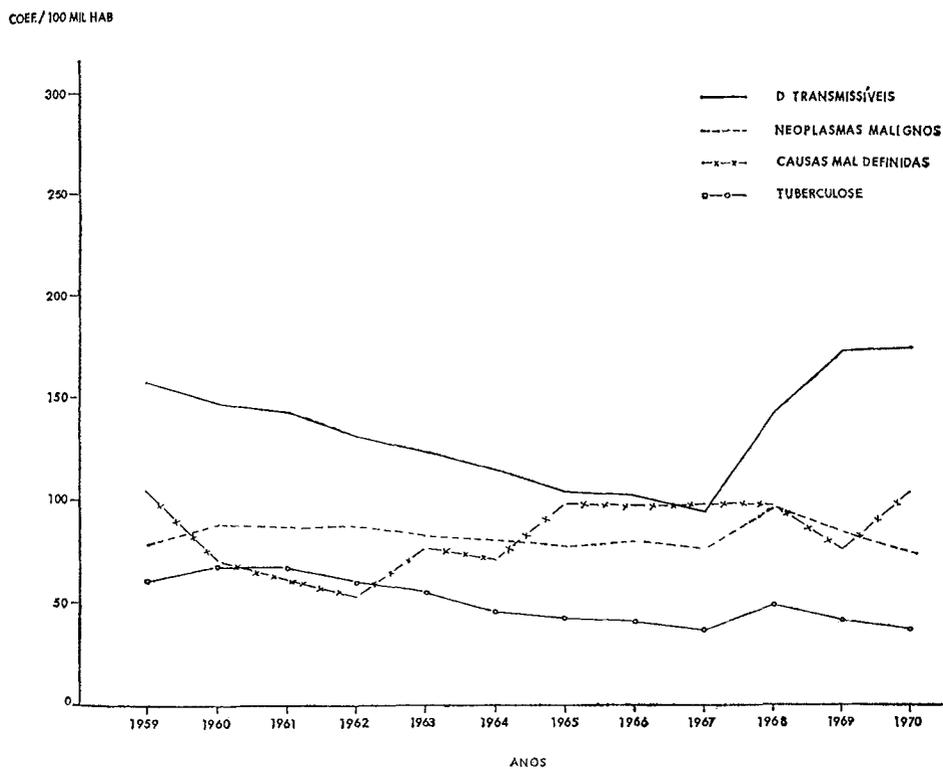


FONTE: 1 — Anuário Estatístico do Brasil (2 — 12)

2 — Séries Estatísticas Retrospectivas — 1970 (14)

GRÁFICO 6

MORTALIDADE POR DOENÇAS TRANSMISSÍVEIS, POR TUBERCULOSE, POR NEOPLASIAS E POR CAUSAS MAL DEFINIDAS (COEFICIENTE POR 100 MIL HABITANTES) — BRASIL — 1959 A 1970

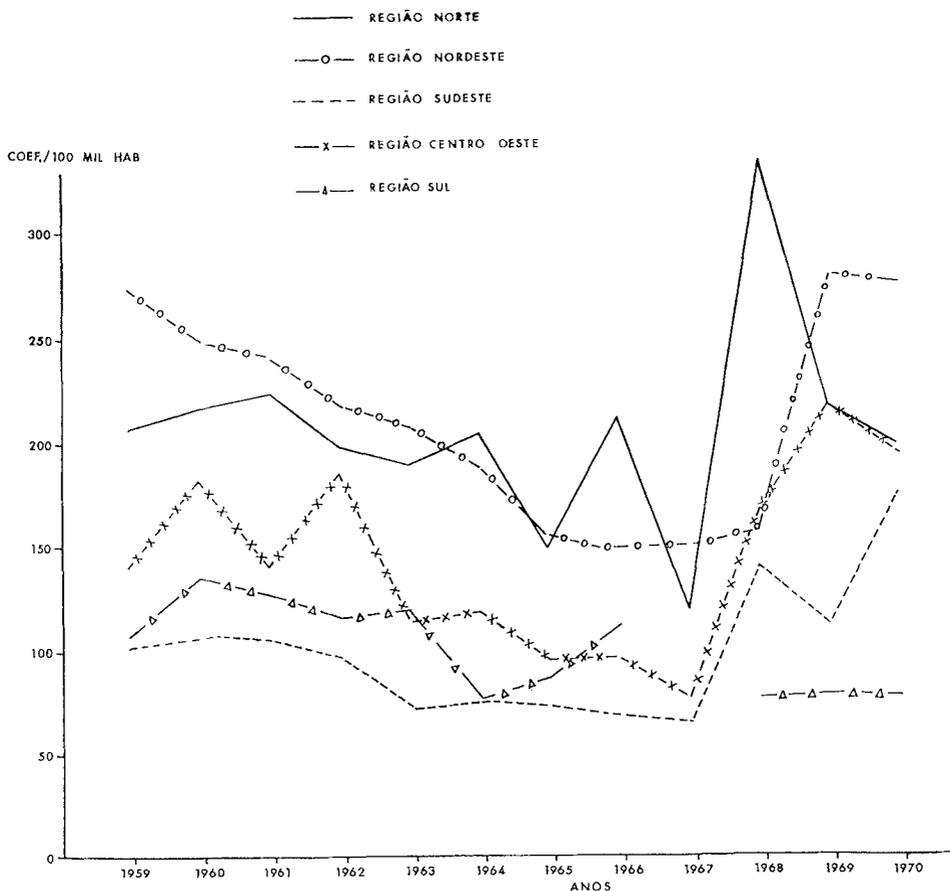


FONTE: 1 — Anuário Estatístico do Brasil (2 — 12)

2 — Séries Estatísticas Retrospectivas — 1970 (14)

GRÁFICO 7

MORTALIDADE POR DOENÇAS TRANSMISSÍVEIS (COEFICIENTE POR 100 MIL HABITANTES) SEGUNDO A REGIÃO FISIOGRAFICA — 1959 A 1970

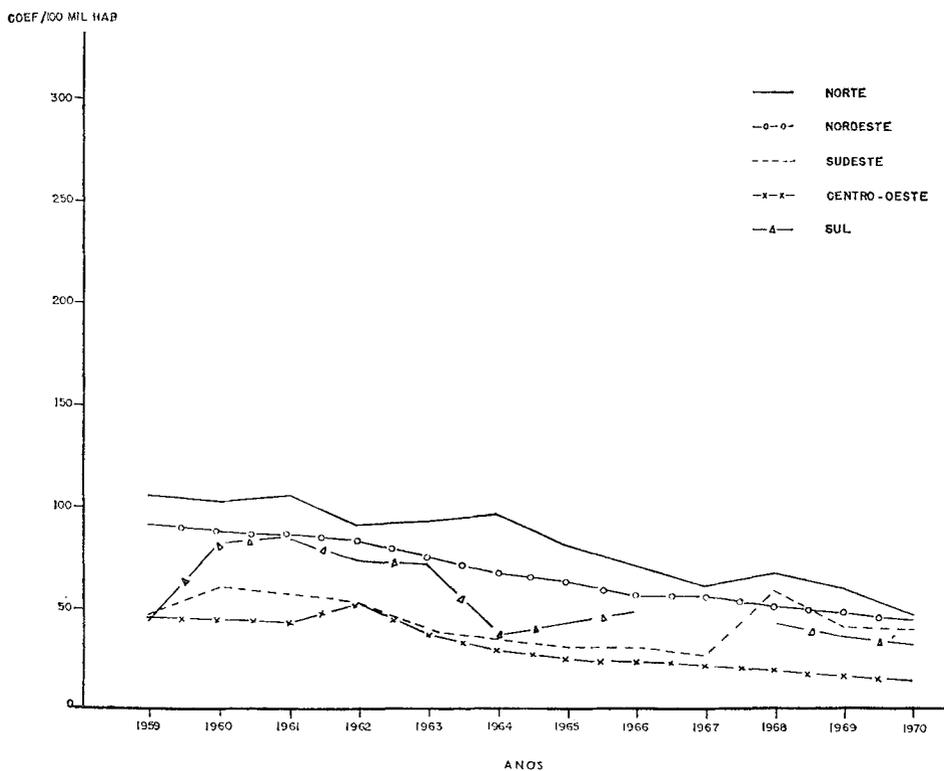


FONTE: 1 — Anuário Estatístico do Brasil (2 — 12)

2 — Séries Estatísticas Retrospectivas — 1970 (14)

GRÁFICO 8

MORTALIDADE POR TUBERCULOSE (COEFICIENTE POR 100 MIL HABITANTES) SEGUNDO A REGIÃO FISIOGRAFICA — 1959 A 1970

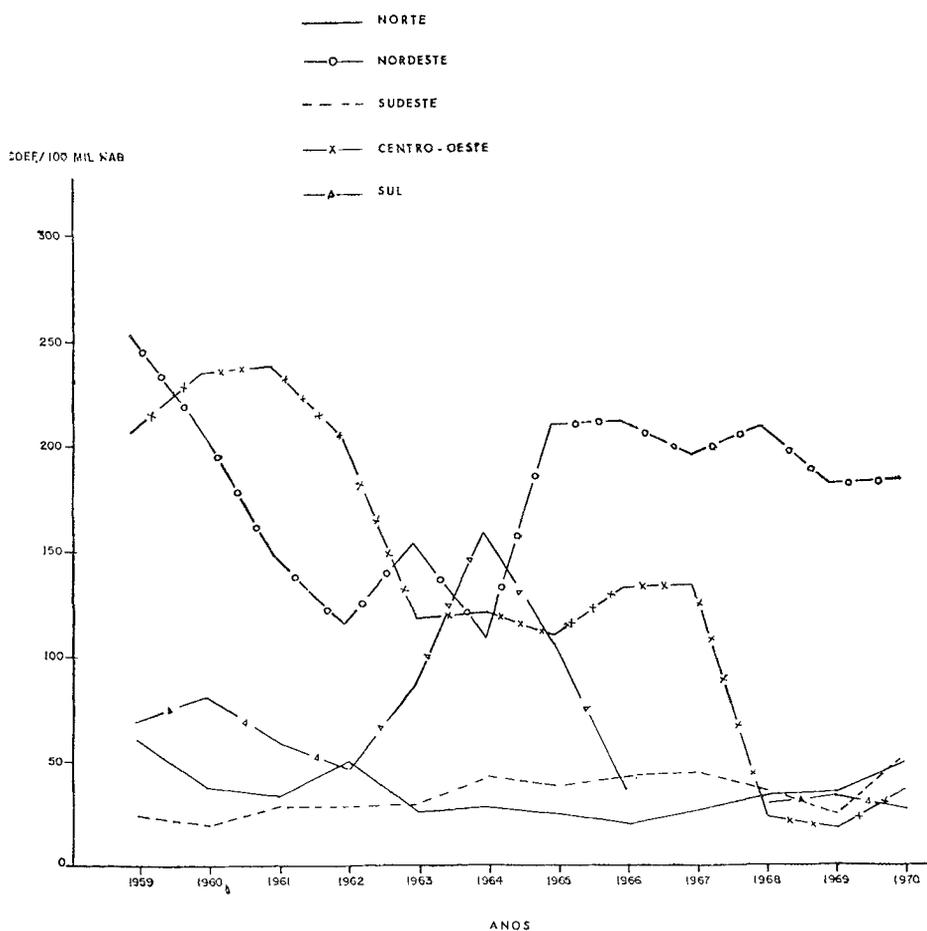


FONTE: 1 — Anuário Estatístico do Brasil (2 — 12)

2 — Séries Estatísticas Retrospectivas — 1970 (14)

GRÁFICO 9

MORTALIDADE POR CAUSAS MAL DEFINIDAS (COEFICIENTE POR 100 MIL HABITANTES) SEGUNDO A REGIÃO FISIOGRAFICA — 1959 A 1970

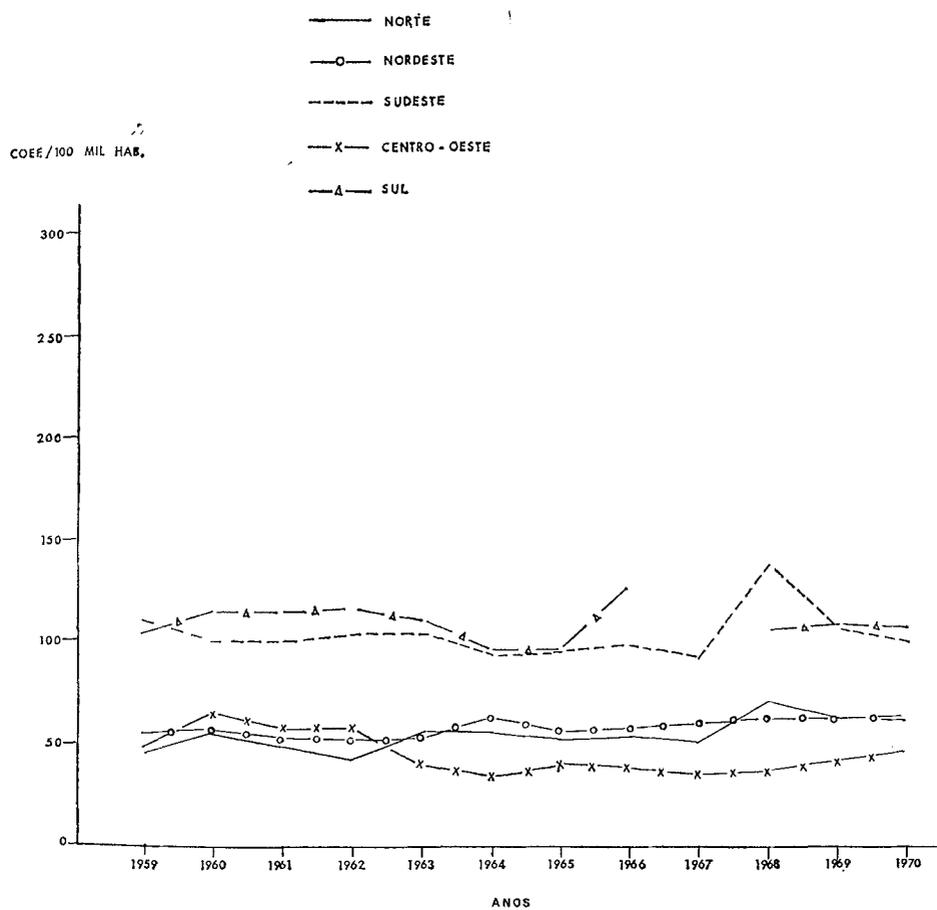


FONTE: 1 — Anuário Estatístico do Brasil (2 — 12)

2 — Séries Estatísticas Retrospectivas — 1970 (14)

GRAFICO 10

MORTALIDADE POR NEOPLASIAS (COEFICIENTE POR 100 MIL HABITANTES) SEGUNDO A REGIÃO FISIOGRAFICA — 1959 A 1970



FONTE: 1 — Anuário Estatístico do Brasil (2 — 12)

2 — Séries Estatísticas Retrospectivas — 1970 (14)

TABELA 1

ÓBITOS POR DOENÇAS TRANSMISSÍVEIS: NÚMERO, PROPORÇÃO SOBRE OS ÓBITOS POR TODAS AS CAUSAS E COEFICIENTE POR 100 MIL HABITANTES, SEGUNDO A REGIÃO FISIOGRAFICA E O ANO

ANO	REGIÃO NORTE			REGIÃO NORDESTE			REGIÃO SUDESTE			REGIÃO CENTRO-OESTE			REGIÃO SU	
	n°	%	Coef	n°	%	Coef	n°	%	Coef	n°	%	Coef	n°	%
1959	1 150	17,3	207,0	7 501	16,0	272,7	11 173	9,9	101,2	289	13,2	141,7	303	3,3
1960	1 240	13,4	216,4	7 104	15,6	248,9	8 508	11,1	107,4	238	16,0	189,8	1 468	11,6
1961	1 333	13,7	222,5	7 138	15,5	239,0	1 722	11,0	105,2	319	14,2	139,9	1 308	11,7
1962	1 242	13,6	197,8	6 821	14,8	213,3	8 432	10,2	97,2	462	16,7	184,3	1 273	11,2
1963	1 243	17,8	189,2	6 647	14,5	205,2	7 960	7,4	71,6	583	11,1	112,1	1 524	11,2
1964	1 412	18,0	205,2	5 089	14,2	186,5	4 376	8,4	74,9	638	13,7	118,5	443	7,5
1965	1 075	16,2	149,4	5 509	12,0	153,3	5 443	8,1	72,0	581	10,6	94,5	595	9,5
1966	1 575	25,7	209,3	5 414	11,8	146,6	4 335	7,5	66,5	639	12,0	94,4	527	9,6
1967	917	15,6	116,4	5 745	12,5	147,9	4 370	7,3	63,4	559	9,9	75,3		
1968	753	30,8	331,8	6 413	13,4	153,1	3 255	11,4	136,8	1 332	24,0	164,0	1 153	9,1
1969	1 855	27,2	214,7	11 786	24,0	277,2	6 561	11,7	110,2	1 703	26,2	214,5	1 165	8,9
1970	1 777	25,1	196,6	12 095	23,8	271,4	3 239	16,9	180,6	1 912	25,6	193,3	1 199	8,7

FONTES: 1 — Anuário Estatístico do Brasil (2 — 12)
2 — Séries Estatísticas Retrospectivas — 1970 (14)

TABELA 2

ÓBITOS POR TUBERCULOSE: NÚMERO, PROPORÇÃO SOBRE OS ÓBITOS POR TODAS AS CAUSAS, PROPORÇÃO SOBRE OS ÓBITOS POR DOENÇAS TRANSMISSÍVEIS E COEFICIENTE POR 100 MIL HABITANTES, SEGUNDO A REGIÃO FISIOGRAFICA E O ANO

ANO	REGIÃO NORTE				REGIÃO NORDESTE				REGIÃO SUDESTE				REGIÃO CENTRO-OESTE				REGIÃO SU		
	n°	% óbitos totais	% óbitos de trans	Coef	n°	% óbitos totais	% óbitos de trans	Coef	n°	% óbitos totais	% óbitos de trans	Coef	n°	% óbitos totais	% óbitos de trans	Coef	n°	% óbitos totais	% óbitos de trans
1959	533	8,8	50,7	105,0	2 505	5,3	33,4	91,0	1 942	4,6	46,5	47,1	94	4,3	32,5	36,1	151	3,8	41,0
1960	583	8,7	47,0	101,7	2 192	5,5	35,1	87,3	4 679	6,1	55,0	59,0	66	3,7	22,9	43,5	881	7,0	60,1
1961	624	8,8	46,8	104,2	2 560	5,6	35,9	85,7	4 711	5,9	51,0	56,3	96	4,3	30,1	42,1	837	7,8	66,8
1962	556	8,2	41,8	89,5	2 563	5,6	37,6	82,0	4 510	5,4	53,5	52,0	131	4,7	28,3	52,4	810	7,1	63,6
1963	601	8,6	49,3	91,5	2 497	5,3	36,2	74,3	2 136	4,0	53,9	38,6	185	3,6	32,8	36,8	908	6,7	59,6
1964	653	8,7	46,2	91,9	1 799	5,0	35,2	65,6	1 952	3,7	41,6	33,4	159	3,3	21,1	28,6	291	3,4	45,4
1965	581	8,7	51,0	89,8	2 182	4,8	39,6	61,5	1 883	3,5	12,4	30,5	149	2,7	25,6	24,2	253	4,1	13,4
1966	522	8,5	33,1	69,3	2 052	4,4	37,7	55,3	1 933	3,4	45,8	30,5	153	3,0	31,7	23,3	226	4,3	44,6
1967	469	6,0	51,1	59,5	2 133	4,6	37,1	51,9	1 805	3,0	41,3	26,2	154	2,7	27,5	20,7			
1968	151	6,2	29,1	66,5	2 036	4,2	31,7	59,2	2 621	4,8	41,9	57,3	152	2,7	11,1	18,7	543	4,3	47,5
1969	593	7,4	27,1	88,2	1 988	4,0	16,9	46,7	2 487	4,3	36,7	49,4	122	1,9	7,2	15,1	517	3,9	41,4
1970	422	6,0	23,7	46,7	1 971	3,9	16,3	41,3	706	3,7	21,8	31,4	113	1,9	7,5	14,8	498	3,6	41,5

FONTES: 1 - Anuário Estatístico do Brasil (2 -- 12)
2 - Séries Estatísticas Retrospectivas (1971 (14)

TABELA 3

ÓBITOS POR NEOPLASIAS: NÚMERO, PROPORÇÃO SOBRE OS ÓBITOS POR TODAS AS CAUSAS E COEFICIENTE POR 100 MIL HABITANTES, SEGUNDO A REGIÃO FISIOGRAFICA E O ANO

ANO	REGIÃO NORTE			REGIÃO NORDESTE			REGIÃO SUDESTE			REGIÃO CENTRO-OESTE			REGIÃO SUL		
	n°	%	Coef	n°	%	Coef	n°	%	Coef.	n°	%	Coef	n°	%	Coef.
1959	245	3,7	44,1	1 473	3,1	53,5	4 543	10,8	110,2	98	4,5	43,0	354	9,0	193,9
1960	318	4,7	55,5	1 556	3,4	54,5	7 893	10,3	99,6	98	5,5	64,5	1 246	9,9	114,2
1961	291	4,1	48,6	1 559	3,4	52,2	8 329	10,5	100,5	129	5,8	56,6	1 196	10,5	114,0
1962	260	3,8	41,4	1 626	3,5	52,0	8 911	10,7	102,8	143	5,2	57,2	1 382	11,3	115,6
1963	359	5,1	54,6	1 715	3,7	53,0	5 716	10,7	103,4	196	3,9	39,0	1 416	10,4	110,5
1964	377	5,0	54,8	1 713	4,8	62,8	5 496	10,5	94,1	187	3,9	33,7	558	9,4	55,7
1965	377	5,7	52,4	1 959	4,3	55,2	5 922	10,9	95,9	242	4,4	39,3	583	9,3	94,0
1966	396	6,4	52,6	2 132	4,6	57,4	6 433	11,2	99,4	243	4,6	36,0	538	11,2	129,4
1967	398	6,8	50,5	2 310	5,0	59,5	6 365	10,7	92,4	252	4,4	34,0			
1968	159	6,5	70,1	2 488	5,2	61,3	6 277	11,5	137,3	297	5,3	36,6	1 605	12,6	103,5
1969	553	8,1	64,0	2 637	5,4	62,0	6 349	11,3	106,6	329	5,1	41,4	1 598	12,1	106,4
1970	554	7,8	61,3	2 838	5,6	63,7	1 773	9,3	99,2	442	5,9	45,8	1 665	12,1	106,6

FONTES: 1 — Anuário Estatístico do Brasil (2 — 12)
2 — Séries Estatísticas Retrospectivas — 1970 (14)

TABELA 4

ÓBITOS POR CAUSAS MAL DEFINIDAS (B45): NÚMERO, PROPORÇÃO SOBRE OS ÓBITOS POR TODAS AS CAUSAS E COEFICIENTE POR 100 MIL HABITANTES SEGUNDO A REGIÃO FISIOGRAFICA E O ANO

ANO	REGIÃO NORTE			REGIÃO NORDESTE			REGIÃO SUDESTE			REGIÃO CENTRO-OESTE			REGIÃO SUL		
	n°	%	Coef	n°	%	Coef	n°	%	Coef	n°	%	Coef	n°	%	Coef
1959	330	5,0	60,0	6 959	14,8	253,0	980	2,3	23,8	424	19,4	207,8	239	6,1	70,2
1960	213	3,1	37,2	5 969	13,1	209,1	1 463	1,9	18,5	357	19,9	235,3	894	7,1	81,9
1961	198	2,8	33,1	4 418	9,6	148,0	2 278	2,9	27,5	544	24,3	233,6	630	5,5	60,0
1962	319	4,7	50,8	3 644	7,9	116,6	2 420	2,9	27,9	514	18,5	205,6	524	4,6	47,2
1963	159	2,3	24,2	5 008	11,0	154,7	1 702	3,2	30,8	602	11,9	119,9	1 113	8,2	86,9
1964	199	2,7	28,9	2 994	3,7	109,7	2 537	4,8	43,4	667	13,8	120,2	935	15,9	160,4
1965	178	2,7	24,8	7 549	16,5	212,9	2 344	4,3	38,0	685	12,4	111,4	715	11,4	115,3
1966	163	2,6	21,6	7 943	17,3	213,9	2 784	4,3	82,7	902	17,0	133,2	172	3,3	37,0
1967	214	3,6	27,2	7 663	16,6	197,3	3 103	5,2	45,0	998	17,6	134,5			
1968	83	3,4	32,7	8 538	17,8	210,5	1 662	8,0	36,3	189	3,4	23,3	511	4,8	23,0
1969	302	3,4	35,0	7 707	15,7	181,3	1 445	8,6	24,3	149	2,3	18,3	504	3,3	33,5
1970	432	6,7	47,8	8 227	16,2	184,6	912	4,8	50,9	350	4,7	35,3	413	3,0	26,4

FONTES: 1 — Anuário Estatístico do Brasil (2 — 12)
2 — Séries Estatísticas Retrospectivas — 1970 (14)

TABELA 5

ÓBITOS, MORTALIDADE PROPORCIONAL E COEFICIENTE POR 100
MIL HABITANTES, SEGUNDO ALGUMAS CAUSAS — BRASIL —
1959 — 1970

ANO	N° de óbitos por doenças transmissíveis	% óbitos totais	Coef. mortal. doenças transmissíveis	N° de óbitos por tuberculose	% óbitos totais	% óbitos doenças transmissíveis	Coef. mortal. tuberculose	N° de óbitos por neopl.	% óbitos totais	Coef. mortal. neopl.	N° de óbitos mal definidos	% óbitos totais
1959	18 400	13,8	137,6	5 275	5,2	39,1	61,6	6 713	6,6	78,4	8 935	8,8
1960	18 606	13,0	147,8	8 701	6,3	46,8	69,1	12 111	7,7	88,2	8 901	6,2
1961	18 8 0	12,9	143,3	8 878	6,1	47,1	67,5	11 504	7,9	87,5	8 068	5,5
1962	18 229	22,1	132,3	8 570	5,7	47,0	62,2	12 222	8,1	88,7	7 421	4,9
1963	13 937	11,2	124,4	6 237	5,0	44,7	53,7	9 402	7,5	83,9	8 584	6,9
1964	11 978	11,3	115,2	4 755	4,5	39,7	45,7	8 331	7,8	80,1	7 332	6,9
1965	12 203	10,3	104,5	5 053	4,2	41,4	43,3	9 083	7,6	77,8	11 471	9,7
1966	12 502	10,4	101,4	4 946	4,1	39,6	40,1	9 842	8,2	79,8	11 964	9,9
1967	11 591	9,9	94,2	4 561	3,9	39,3	37,1	9 325	7,9	75,8	11 978	10,9
1968	15 906	12,9	141,5	5 508	4,5	34,6	49,0	10 826	8,8	96,3	10 983	8,9
1969	23 070	17,5	172,6	5 537	4,2	24,0	41,4	11 466	8,7	83,8	10 107	7,7
1970	20 222	29,6	208,0	3 743	3,8	18,9	38,7	7 277	7,4	75,2	10 334	10,5

FONTE: 1 — Anuário Estatístico do Brasil (2 — 12)

2 — Séries Estatísticas Retrospectivas — 1970 (14)

APLICAÇÃO DE MODELOS MULTIDIMENSIONAIS À AGRICULTURA

Leila Maia *

do Instituto de Pesquisas do IPEA

SUMÁRIO

1. *Introdução*
2. *Culturas Permanentes: Um estudo de caso, o café*
3. *Culturas Temporárias: Um Estudo Geral*
4. *O Dual*
5. *Apreciação Final*

1. INTRODUÇÃO

Dentre as principais aplicações da programação matemática, em particular a linear, destacam-se os modelos de produção e distribuição na agricultura. Basicamente dois fatos explicam sua aplicabilidade a este tipo de problema. A princípio observa-se que os mercados para pro-

* A autora agradece aos economistas Claudio Contador e Hamilton C. Tolosa pelas críticas e comentários à primeira versão deste trabalho.

dados agrícolas se aproximam bastante das hipóteses de concorrência perfeita:

- a) **Atomicidade.** O mercado para produtos agrícolas geralmente compreende um grande número de produtores e compradores com diminuta influência no preço final do produto.
- b) **Homogeneidade.** As possíveis diferenciações no produto final podem ser consideradas negligíveis.
- c) **Grau de informação sobre o mercado.** A existência de Bolsas para os produtos agrícolas contribuem para a fluidez da informação sobre as condições do mercado.

Outra razão importante é que a função de produção homogênea do primeiro grau apresenta boa aderência estatística à maioria das atividades agrícolas. Em estudos para curto prazo esta hipótese de rendimentos constantes de escala se aproxima da realidade, o que significa dizer que, se multiplicarmos o vetor de insumos (x) por um escalar (λ), o produto (y) aparecerá multiplicado por este mesmo escalar, isto é, se

$$y = f(x)$$

então

$$\lambda y = f(\lambda x)$$

É fato conhecido que a solução da programação linear é consistente com a concorrência perfeita. Na solução ótima o preço do produto é igual ao custo marginal para as atividades com produção positiva (atividades básicas) e menor do que o custo marginal para as atividades com produção igual a zero (atividades não básicas).

As aplicações da programação linear à agricultura têm por tradição dado ênfase à escolha entre tecnologias e produtos alternativos. A função objetivo usual procura a maximização da renda líquida recebida pelos agricultores sujeita a restrições lineares impostas pela matriz tecnológica ou de coeficientes técnicos e pela disponibilidade dos fatores de produção.

O presente artigo descreve um tipo menos comum de modelo agrícola, ou seja: extensões do modelo de transportes (também denominado de Hitchcock-Koopmans) para a determinação de níveis eficientes de produção e distribuição de um ou mais produtos agrícolas entre diferentes áreas ou centros de produção, armazenagem e consumo. Essas

aplicações implicam em ampliar a versão original do modelo de transportes¹ em duas direções:

a) Introduzindo novas restrições além daquelas sobre a capacidade de produção e demanda. É o caso, por exemplo, de limites superiores impostos às quantidades transportadas com origem em determinado centro de produção.

b) Empregando modelos multidimensionais, onde o número de dimensões ou índices pode ser tão grande quanto desejado. A formulação teórica desta classe de modelos constitui o interesse central do presente artigo. Note-se, entretanto, que embora os modelos multidimensionais de produção e distribuição possam ser facilmente ajustados às necessidades práticas, isto acarreta um acréscimo proporcional de complexidade, e altos custos para a solução e simulação do modelo.

Como exemplos de variáveis multidimensionais de complexidade crescente temos:

x_{ij} = número de unidades de um produto agrícola transportadas da região de produção i , para à região de consumo j

x_{ikj} = número de unidades de um produto agrícola transportadas do centro de produção i , para o centro de consumo j , via centro de armazenagem k

x_{ikjt} = número de unidades de um produto agrícola transportadas do centro de produção i , para o centro de consumo j , via centro de armazenagem k , durante o período t

x_{gikjt} = número de unidades do produto agrícola g , transportadas do centro de produção i , para o centro de consumo j , via centro de armazenagem k , durante o período t

As duas últimas variáveis acima introduzem as dimensões tempo e múltiplos produtos no modelo. No presente estudo, nos limitaremos ao caso onde $g = 1$ e t constante, isto é, ao modelo estático com um único produto e vários centros de produção (i), de consumo (j) e armazenagem (k). Convém ressaltar, contudo, que várias das conclusões obtidas para três dimensões podem ser estendidas para o caso de múltiplos produtos, meios alternativos de transporte e restrições sobre as quantidades transportadas entre dois centros quaisquer. A generaliza-

¹ O modelo original de transportes pode ser escrito numa forma matricial e compacta como

Min cx

Sujeita a $Ax = \begin{bmatrix} s \\ d \end{bmatrix} \quad x \geq 0$

Onde c é o vetor linha dos custos de transportes, s e d subvetores da capacidade de produção e demanda, x o vetor de quantidades transportadas e A uma matriz binária, isto é, composta de elementos iguais a zero e a um

ção para a versão dinâmica torna-se mais problemática uma vez que envolve a especificação de equações para o aumento da capacidade de produção em cada um dos centros de produção ².

Após ter apresentado as principais características dos modelos de produção e distribuição, voltamos ao caso de três dimensões. Já foi mencionado que o objetivo principal deste artigo é mostrar de que modo estes modelos podem ser aplicados à agricultura diferenciando o caso de culturas permanentes ou temporárias. Define-se neste trabalho cultura permanente como sendo aquela em que de um só plantio podem ser extraídas várias colheitas, sem que a plantação seja destruída no processo, como por exemplo o café e o algodão. Por outro lado, a cultura temporária será sempre destruída como consequência da colheita, o que implicará em novo processo de plantação para que se obtenha outra safra, como, por exemplo a da soja e a do trigo.

A estrutura básica do modelo ³ a ser discutido possui o seguinte formato:

$$\text{Min } F(x) = \sum_{i=1}^m \sum_{k=1}^p \sum_{j=1}^n c_{ikj} x_{ikj}$$

Sujeita às restrições:

i) Restrição sobre a capacidade de produção

$$\sum_{k=1}^p \sum_{j=1}^n x_{ikj} \leq s_i \quad i = \dots m; s_i > 0$$

ii) Restrição sobre a capacidade de armazenagem

$$\sum_{i=1}^m \sum_{j=1}^n x_{ikj} \leq a_k \quad k = 1 \dots p; a_k > 0$$

iii) Restrição sobre a demanda

$$\sum_{i=1}^m \sum_{k=1}^p x_{ikj} \geq d_j \quad j = 1 \dots n; d_j > 0$$

² Para uma aplicação dinâmica à siderurgia brasileira, ver, por exemplo, KENDRICK, D A *Programming Investment in the Processing Industries An Approach to Sectoral Planning*. Cambridge, MIT Press, 1967

³ Existem variações do modelo tridimensional dentre as quais destaca-se o chamado "Solid Transportation" definido para vários produtos, centros de produção e consumo. Neste caso, a variável quantidade transportada é escrita como $x_{\theta ij}$. Há, além das desigualdades de capacidade de produção e demanda, uma terceira restrição definida como

$$\sum_{\theta} x_{\theta ij} = z_{ij}$$

isto é, o somatório do fluxo de todas as remessas entre i e j deve ser igual à constante z_{ij} . Foram desenvolvidos algoritmos específicos para essa classe de modelos, ver, por exemplo, HALEY, K B. *The Solid Transportation Problem Operations Research*, jul/ago 1962

iv) Condições de equilíbrio no mercado

$$\sum_{i=1}^m s_i = \sum_{k=1}^p a_k = \sum_{j=1}^n d_j$$

v) Restrição de não-negatividade

$$x_{ikj} > 0 \quad i = 1 \dots m; \quad j = 1 \dots n \quad c_{ikj} \geq 0 \\ k = 1 \dots p$$

Num total de $(m + p + n + 1)$ restrições, excluídas as de não-negatividade. Os parâmetros do modelo têm a seguinte definição:

c_{ikj} = custos unitários de produção, armazenagem e distribuição de um determinado produto agrícola com origem no centro de produção i , destino no centro de consumo j e via centro de armazenagem k . Pressupõe-se que a função de produção opere com rendimentos constantes de escala e que os custos de transportes sejam do tipo proporcional ou uniforme

s_i = capacidade de produção do centro i

a_k = capacidade de armazenagem no centro k

d_j = demanda no centro de consumo j

Em síntese, o modelo acima é aplicado a um período curto de tempo, geralmente um ano, e consiste na determinação de níveis de produção e distribuição de um dado produto agrícola (x_{ikj}) que, satisfeitas as restrições de produção, armazenagem e demanda, minimizam a função de custo total de produção, armazenagem e transporte ⁴.

2. CULTURAS PERMANENTES: UM ESTUDO DE CASO, O CAFÉ

Teoricamente o modelo apresentado na seção anterior pode refletir qualquer situação encontrada na produção e distribuição dos produtos agrícolas. A hipótese de culturas permanentes aceita como fixa a produção total em cada região, tornando o modelo consistente com a ocorrência de superprodução nas regiões produtoras e capacidade ociosa nos centros de armazenagem.

Para melhor ilustrar a aplicação dos modelos multidimensionais, consideremos como estudo de caso, a produção e distribuição de café no Brasil. Esta escolha foi feita por serem as características da lavoura do café coincidentes com a definição adotada para culturas permanen-

⁴ De modo a preservar a condição de linearidade do modelo, admite-se que a função do custo total é aditiva e separável

tes; um cafeeiro exige de três a cinco anos para se desenvolver e começar a dar frutos, e pode manter-se comercialmente produtivo até a idade de trinta anos ou mais. A política brasileira de café tem se caracterizado por forte intervenção no mercado interno através de preços mínimos, confisco cambial, controle de estoques, preços controlados para o consumo interno, etc. A produção tem apresentado períodos de superprodução e acumulação de estoques alternados com épocas de escassez em virtude principalmente de fatores climáticos (geadas). Nos últimos cinco anos, a produção média tem se mantido em torno de 19,0 milhões de sacas, contra 18,8 milhões de exportação e 8,1 milhões de consumo interno.

No caso concreto, propõe-se como ponto de partida a existência de:

- i) sete zonas produtoras, isto é, $i = 1 \dots 7$, com diferentes características em termos de utilização da mão-de-obra, fertilizantes, maquinaria e tipo de solo; s_i é a capacidade de produção média anual em milhões de sacas na zona i .
- ii) três centros de armazenagem, $k = 1 \dots 3$, localizados no interior, fora das zonas produtoras e dos centros de consumo, a_k é a capacidade de armazenagem média anual em cada um destes centros em milhões de sacas.
- iii) três centros de consumo, $j = 1 \dots 3$, que no presente caso se confundem com os três principais portos para exportação do café⁵. O consumo interno poderia ser alocado a esses três portos de acordo com um critério de demanda por região brasileira. Nessas condições, a quantidade demandada em cada centro de consumo (d_j) será igual à média do café exportado por porto. Se o interesse estiver concentrado exclusivamente na exportação supõe-se que o consumo interno seja igual a zero⁶.

Em resumo: numa versão simplificada supõe-se que $i = 1 \dots 7$, $k = 1 \dots 3$ e $j = 1 \dots 3$. Uma aproximação mais realista exigiria mais de vinte zonas produtoras e um número maior de centros de armazenagem e portos.

A discussão do modelo distingue três estágios, entre os quais a diferença marcante está na maneira como é tratada a localização dos centros de armazenagem, isto é, estejam eles localizados apenas no interior, ou coincidam com as regiões de produção ou consumo ou ambas.

5 Nesta etapa as zonas produtoras poderiam ser Norte do Paraná, Sorocabana, Nordeste, Araguaquense, Mogiana, Sul de Minas Gerais e Espírito Santo. Os centros de armazenagem no interior seriam: Londrina, Campinas e Juiz de Fora; e os portos: Rio de Janeiro, Santos e Paranaguá.

6 É evidente que, nesse caso, o consumo interno deveria ser subtraído da produção total (Σs_i) e da capacidade total de armazenagem (Σa_k).

Para qualquer dos estágios verificam-se os seguintes postulados:

- Admite-se que não haja entre as zonas de produção diferenças relevantes quanto à clima e estrutura fundiária.
- Os custos de produção e distribuição diferem entre as regiões mas são constantes a curto prazo⁷.
- Os custos de distribuição internos a cada região são negligenciáveis.
- Cada remessa de uma zona produtora para um porto passa por um e apenas um armazém.

Além disto, a estrutura de custos ou seja, c_{ikj} pode ser decomposta da seguinte maneira

$$c_{ikj} = e_i + t_{ikj}$$

onde

e_i = custos unitários de produção na região i

t_{ikj} = custos unitários de distribuição de i até j , via k

Para os custos de produção tem-se que:

$$e_i = e_{i1} + e_{i2}$$

onde e_{i1} é o custo com a mão-de-obra, maquinaria, fertilizantes e defensivos usados na manutenção da plantação e colheita do café, enquanto e_{i2} representa os custos com a mão-de-obra e maquinaria para seleção, secagem e ensacamento do café. É conveniente distinguir entre estas duas componentes da função de custo de produção pois representam fases no tratamento do café que ocorrem em épocas distintas e com intensidades diversas, principalmente em termos de utilização da mão-de-obra. As duas parcelas de custo são expressas em sacos ou toneladas de café por hectare.

Os custos de distribuição podem ser desdobrados em:

$$t_{ikj} = t_{ik} + t_{kj}$$

ou ainda:

$$t_{ik} = r_k + T_{ik}$$

$$t_{kj} = r_j + T_{kj}$$

⁷ Na realidade, aqui faz-se também abstração do problema da flutuação nos preços internacionais e preços mínimos o que significa dizer que para efeito do modelo estes preços são dados e constantes a curto prazo. A demanda pelas exportações brasileiras poderia ser definida como $D = D(p, q, P)$ onde p é a cotação do café Santos 4 na bolsa de Nova York, q é a quantidade exportada e P um vetor de preços dos principais cafés, arábica e robusta (africano), concorrentes do Brasil. Em condições normais a quantidade produzida (Σs_p) seria elástica, com uma defasagem média de três anos, ao preço internacional. Na hipótese de existirem estoques reguladores, a elasticidade é instantânea.

onde r_k e r_j são respectivamente os custos unitários de armazenagem em k e em j , enquanto que T_{ik} e T_{kj} representam os custos de transporte entre i, k e j .

Quando se permitir armazenagem nas regiões produtoras (segundo e terceiro estágio), os custos destes serviços serão incluídos no custo de produção. Da mesma forma, quando existirem diferenças importantes entre taxas e impostos estas podem ser incorporadas aos parâmetros de custo.

2.1 — Primeiro estágio: Centro de armazenamento no interior

Voltando ao modelo no seu primeiro estágio pode-se escrever que:

$$\text{Min } F(x) = \sum_{i=1}^{\gamma} \sum_{k=1}^{\mathcal{S}} \sum_{j=1}^{\mathcal{S}} c_{ikj} x_{ikj}$$

Sujeito às restrições:

$$\sum_{k=1}^{\mathcal{S}} \sum_{j=1}^{\mathcal{S}} x_{ikj} \leq s_i \quad i = 1 \dots \gamma$$

$$\sum_{i=1}^{\gamma} \sum_{j=1}^{\mathcal{S}} x_{ikj} \leq a_k \quad k = 1 \dots \mathcal{S}$$

$$\sum_{i=1}^{\gamma} \sum_{k=1}^{\mathcal{S}} x_{ikj} \geq d_j \quad j = 1 \dots \mathcal{S}$$

com as condições de equilíbrio e não-negatividade

$$\sum_{i=1}^{\gamma} s_i \geq \sum_{k=1}^{\mathcal{S}} a_k = \sum_{j=1}^{\mathcal{S}} d_j$$

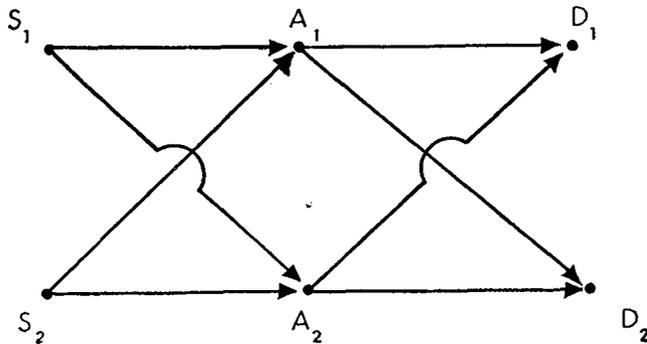
$$X_{ikj} \geq 0$$

Neste ponto, é oportuno que se faça algumas observações a propósito do modelo acima. Em relação às componentes do parâmetro custos, tem-se que $r_j = 0$. Além disto, verifica-se que a condição de equilíbrio foi relaxada, de modo a permitir excesso de produção. Para a solução do problema, introduz-se um vetor de variáveis de folga nas equações das zonas produtoras, sendo que o valor destas variáveis na solução ótima indica como aquele excesso se distribui pelas regiões produtoras. Na versão dinâmica do modelo o excesso é transferido para o ano seguinte e somado à produção corrente daquele ano.

Admitindo que a condição de equilíbrio entre produção e armazenagem seja satisfeita, o primeiro estágio do modelo compreende 12 restrições (exclusive a de não-negatividade) em 63 variáveis. Genericamente, o caso de m regiões produtoras, p centros de armazenagem e n portos ter-se-ia $(m + p + n - 1)$ restrições linearmente independentes

em $m \cdot n \cdot p$ variáveis, significando que o vetor da solução ótima contém $(m + p + n - 1)$ atividades positivas ($m \cdot p \cdot n - m - p - n + 1$) atividades iguais a zero.

Ao modelo de produção e distribuição pode ser associado um grafo orientado. Essa forma possui a vantagem de permitir a melhor visualização das rotas alternativas entre os vértices (ou centros) de produção, armazenagem e consumo. Num exemplo simplificado do primeiro estágio, com duas zonas de produção, (S_1 e S_2), dois centros de armazenagem (A_1 e A_2) e dois centros de consumo, (D_1 e D_2), a aparência do grafo será:



Ao grafo acima corresponde uma matriz binária ou de conexidade onde os valores 1 indicam remessa positiva e zero um fluxo igual a zero ou seja:

$$\begin{array}{c}
 \begin{matrix}
 S_1 \\
 S_2 \\
 A_1 \\
 A_2 \\
 D_1 \\
 D_2
 \end{matrix}
 \begin{bmatrix}
 S_1 & S_2 & A_1 & A_2 & D_1 & D_2 \\
 0 & 0 & 1 & 1 & 0 & 0 \\
 0 & 0 & 1 & 1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 1 & 1 \\
 0 & 0 & 0 & 0 & 1 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0
 \end{bmatrix}
 \end{array}$$

Em relação ao modelo, cada variável x_{ijk} corresponde à conjunção de dois vetores. Ex: x_{121} é representada pelos vetores

$$\overrightarrow{S_1 A_2} \text{ e } \overrightarrow{A_2 D_1}$$

2.2 — Segundo estágio:

Neste estágio a principal modificação é que agora admite-se armazéns localizados não apenas no interior mas dois casos alternativos que são:

— os armazéns estão localizados no interior e também nas regiões produtoras, dando aos agricultores as opções de estocar o café na própria área ou enviá-los para os centros de armazenagem no interior.

— os armazéns estão localizados no interior e também nos portos, dando aos agricultores a alternativa de enviar o café para armazenagem no interior ou nos portos.

Os dois casos vistos podem ser tratados de maneira similar, e por isto nos limitaremos ao primeiro. Neste estágio, o número de armazéns é aumentado para 10 (soma dos centros de armazenagem no interior e das zonas produtoras). A forma analítica é a mesma do primeiro estágio, com a diferença que o somatório em k varia de 1 a 10, e introduz-se sete novas equações para os centros de armazenagem localizados nas zonas de produção. Este segundo estágio possui 19 equações linearmente independentes em 10 atividades ⁸.

A matriz do exemplo anterior assume agora o seguinte aspecto:

$$\begin{array}{cccccc}
 & S_1 & S_2 & A_1 & A_2 & D_1 & D_2 \\
 S_1 & \left[\begin{array}{cccccc}
 1 & 1 & 1 & 1 & 1 & 1 \\
 1 & 1 & 1 & 1 & 1 & 1 \\
 0 & 0 & 0 & 0 & 1 & 1 \\
 0 & 0 & 0 & 0 & 1 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0
 \end{array} \right] \\
 S_2 & & & & & & \\
 A_1 & & & & & & \\
 A_2 & & & & & & \\
 D_1 & & & & & & \\
 D_2 & & & & & &
 \end{array}$$

Os vetores $\overrightarrow{S_1 S_1}$ e $\overrightarrow{S_2 S_2}$ significam um “loop” no grafo, e implicam na possibilidade de armazenagem do café na própria região de produção. Por sua vez, $\overrightarrow{S_1 S_2}$ e $\overrightarrow{S_2 S_1}$ representam a mesma ligação porém em sentidos opostos. Na prática, dificilmente ambas seriam incluídas simultaneamente na solução ótima.

No que concerne aos custos unitários correspondentes, estes podem ser definidos como ⁹.

$$c_{ikj} = e_i + t_{ik} + t_{hj} = e_i + r_k + T_{ik} + T_{kj} \text{ para } i \neq k \neq j$$

$$c_{ij} = e_i + r_i + T_{ij} \text{ para } i = k \neq j$$

2.3 — Terceiro estágio:

Neste estágio admite-se a localização dos armazéns nas zonas produtoras, no interior e nos portos. O agricultor tem a possibilidade de estocar a produção em qualquer desses centros antes de exportá-la.

⁸ Genericamente: Se $i = 1 \dots m$, $k = 1 \dots (p+m)$; $j = 1 \dots n$, então tem-se $n + p + 2m - 1$ equações linearmente independentes em $m \cdot n \cdot (p+m)$ atividades, onde cada atividade é definida para:

$$x_{ikj} \geq 0 \text{ quando } \begin{cases} i \neq k \neq j \\ i = k \neq j \end{cases}$$

⁹ Note-se que, por hipótese, $T_{ii} = 0$

A diferença básica entre este estágio e os anteriores é que aqui admite-se simultaneamente a existência de armazéns nas regiões de produção, no interior e nos portos. Neste caso, teríamos 7 zonas produtoras, 13 centros de armazenagem e 3 centros de consumo, abrindo as possibilidades de armazenagem já descritas. Há, portanto, 22 equações linearmente independentes em 273 atividades. Generalizando: $2(m+n)+p-1$ relações independentes em $m.n(m+n+p)$ variáveis. E sendo assim, como nas seções anteriores a matriz de conexidade toma a forma:

$$\begin{array}{cccccc}
 & S_1 & S_2 & A_1 & A_2 & D_1 & D_2 \\
 \begin{array}{l} S_1 \\ S_2 \\ A_1 \\ A_2 \\ D_1 \\ D_2 \end{array} & \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}
 \end{array}$$

Foram acrescentadas, em relação ao estágio anterior, duas novas possibilidades de "loop" nos portos, indicadas por $\overrightarrow{D_1 D_1}$ e $\overrightarrow{D_2 D_2}$ e remessas em sentido contrário, $\overleftarrow{D_1 D_2}$ e $\overleftarrow{D_2 D_1}$. Está claro que como no caso anterior, dificilmente os dois últimos vetores estarão presentes simultaneamente na solução ótima.

O terceiro estágio representa o caso mais geral do modelo em discussão, é portanto conveniente que se examine os diferentes aspectos do parâmetro de custos em cada situação oferecida ao agricultor. Na prática, ter-se-á cinco possibilidades, apresentadas no quadro a seguir:

ALTERNATIVA	ATIVIDADE (X_{ikj})	PARÂMETRO DE CUSTO	OBSERVAÇÕES
1	$i = k \neq j$	$C_{ij} = e_i + r_i + T_{ij}$	Armazenagem na própria zona produtora
2	$i \neq i^* \neq j$	$C_{ii^*} = e_i + r_{i^*} + T_{ii^*} T_{i^*j}$	Armazenagem em outra zona de produção (i^*)
3	$i \neq k \neq j$	$C_{ikj} = e_i + r_k + T_{ik} + T_{kj}$	Armazenagem no interior
4	$i \neq k = j$	$C_{ijj} = e_i + r_j + T_{ij}$	Armazenagem no próprio porto de exportação
5	$i \neq j \neq j^*$	$C_{ijj^*} = e_i + r_{j^*} + T_{ij} + T_{jj^*}$	Armazenagem em outro porto (j^*) que não o de exportação

A segunda e quinta alternativas são menos prováveis de ocorrer na prática.

3. CULTURAS TEMPORÁRIAS: UM ESTUDO GERAL

Para as culturas permanentes, a produção em cada região é tomada como dado do problema. A aplicação desse mesmo modelo às culturas temporárias, como, por exemplo, a soja, implica numa importante consideração adicional: a decisão acerca da quantidade de cada fator de produção (como terra, trabalho, capital) a ser utilizada em cada período. Por isto, a quantidade produzida por uma região aparece como variável no modelo, correspondendo à soma de todas as remessas originais da região em questão. Essas remessas, por outro lado, são limitadas pela disponibilidade de cada fator de produção dentro da região.

O modelo abaixo apresenta, a título de ilustração, um exemplo com dois fatores limitativos: a terra e o trabalho¹⁰. A terra é analisada como um fator inteiramente imóvel entre regiões, enquanto a mão-de-obra é tratada como um recurso com mobilidade perfeita, via migrações entre as várias regiões produtoras¹¹.

A fim de simplificar, a análise será limitada ao primeiro estágio do modelo, isto é, a armazenagem só é permitida nos centros do interior. Suponha-se, como na fase anterior, que existam m zonas produtoras, p centros de armazenagem e n portos. A mão-de-obra é tida como fator de produção homogêneo para todas as regiões, e ao mesmo tempo, a terra pode ser diferenciada entre as regiões, embora seja homogênea internamente a cada uma delas.

Nessas condições, pode-se escrever a função objetivo como:

$$\text{Min } F(x) = \sum_{i=1}^m \sum_{k=1}^p \sum_{j=1}^n c_{ikj} x_{ikj}$$

Sujeita às seguintes restrições:

i) Restrições sobre a capacidade de produção

$$g_i \sum_{k=1}^p \sum_{j=1}^n x_{ikj} \leq L_i \quad i = 1 \dots m; L_i > 0$$

$$\sum_{i=1}^m \sum_{k=1}^p \sum_{j=1}^n h_i x_{ikj} \leq M; \quad M > 0$$

onde g_i e L_i são respectivamente o coeficiente técnico para a terra e a quantidade disponível de terra em cada região produtora i . Por seu turno, h_i e M correspondem ao coeficiente técnico do trabalho na re-

10 Em princípio pode-se de igual modo especificar uma ou mais restrições sobre o capital. Isto, entretanto, acarreta dificuldades relacionadas à hipótese de mobilidade entre as regiões produtoras. O fator capital seria considerado imóvel no caso das máquinas, equipamentos e instalações, e suposto como móvel entre as regiões no caso do capital de giro e fluxos financeiros.

11 Considera-se aqui apenas a mão-de-obra diretamente ligada à produção. Supõe-se que não ocorram limitações com respeito, por exemplo, à mão-de-obra empregada nos serviços de armazenagem e operação nos portos.

gião i e a disponibilidade da mão-de-obra para o conjunto das regiões produtoras¹².

ii) Restrições sobre a capacidade de armazenagem

$$\sum_{i=1}^m \sum_{j=1}^n x_{ikj} \leq a_k \quad k = 1 \quad p; a_k > 0$$

iii) Restrições sobre a demanda

$$\sum_{i=1}^m \sum_{k=1}^p x_{ikj} \geq d_j \quad j = 1 \dots n, d_j > 0$$

iv) Restrições de equilíbrio no mercado e não negatividade

$$\sum_{k=1}^p a_k = \sum_{j=1}^n d_j$$

$$x_{ikj} \geq 0$$

Cabem aqui alguns comentários a propósito das restrições do modelo. De início observa-se pelas restrições sobre a capacidade de produção que a disponibilidade do fator terra é fixada para cada região, ao mesmo tempo que, para mão-de-obra a restrição refere-se ao conjunto das regiões. Verifica-se, ainda, pela restrição de equilíbrio do mercado que, dada a demanda, a ocorrência de superprodução é literalmente afastada. Entretanto, pode acontecer da quantidade produzida não ser suficiente para absorver totalmente o fator terra e/ou mão-de-obra. Nesse caso o fator (ou fatores) ficará subutilizado.

A restrição de equilíbrio pode entretanto ser relaxada, de tal maneira que se permita capacidade ociosa nos centros de armazenagem, isto é:

$$\sum_k a_k \geq \sum_j d_j$$

nessa situação, pode-se impor ao sistema o pagamento de uma penalidade ou multa por unidade de armazenagem não utilizada. Sendo q_k o valor desta penalidade, a função objetivo pode ser reescrita da seguinte maneira:

$$\text{Min } F(x) = \sum_{i=1}^m \sum_{k=1}^p \sum_{j=1}^n c_{ilkj} x_{ilkj} + \sum_{k=1}^p \left(a_k - \sum_{i=1}^m \sum_{j=1}^n x_{ilkj} \right) q_k$$

ou seja, minimizar a soma dos custos de produção, distribuição e multas pagas pela ociosidade dos armazéns.

¹² Se L_i é medido em hectares e x_{ikj} em toneladas, a unidade de q_k será ton/hectare. Da mesma maneira, se M é medido em homens/dia, a unidade de h_i será homens/dia/ton.

4. O DUAL

A alocação ótima de recursos num modelo multidimensional pode ser vista sob duas óticas alternativas: uma como um problema de determinação de remessas que minimizam a função de custo total, outra como um problema de imputar preços aos fatores empregados no processo de produção e distribuição dos produtos agrícolas. O primeiro problema corresponde ao problema primal discutido nas seções anteriores, o segundo diz respeito ao seu dual.

Para iniciar, tome-se o dual do primeiro estágio do modelo para culturas permanentes. Suponha-se, em seguida, que as duas primeiras restrições do modelo primal são multiplicadas por -1 . Isto feito, o problema dual terá a forma:

$$\text{Max } D(v, z, u) = \sum_{j=1}^n d_j v_j + \sum_{k=1}^p -a_k z_k + \sum_{i=1}^m -s_i u_i$$

sujeita às restrições:

$$i = 1 \quad m, \quad k = 1 \quad p, \quad j = 1 \quad n$$

$$v_j - z_k - u_i \leq c_{ikj}$$

$$v_j \geq 0; \quad z_k \geq 0; \quad u_i \geq 0$$

O problema dual presta-se a uma interpretação bastante interessante do ponto de vista econômico. Pelo teorema da dualidade sabe-se que, na solução ótima ¹³

$$F(X^*) = D(v^*, z^*, u^*)$$

onde X^* , v^* , z^* , u^* são respectivamente os vetores de atividade e preços duais na solução ótima.

Então:

$$\frac{\partial F(X^*)}{\partial s_i} = u_i^*; \quad \frac{\partial F(X^*)}{\partial a_k} = z_k^*; \quad \frac{\partial F(X^*)}{\partial d_j} = v_j^*$$

Das duas primeiras derivadas interpreta-se imediatamente u_i^* e z_k^* como custos marginais de produção e armazenagem. Para interpretar v_j^* , comecemos por somar uma constante, p ao termo à esquerda da restrição dual, com o que obtém-se:

$$(v_j + p) - (u_i + z_k + p) \leq c_{ikj}$$

¹³ Indicada por um asterisco

o valor de p é definido como preço base do produto agrícola em questão para qualquer centro, seja de produção, armazenagem ou consumo. Com mais precisão: p compreende custos unitários de produção e armazenagem em cada centro, pode-se então escrever que:

$$p = e_i + r_k$$

donde:

$$(v_j + e_i + r_k) - (u_i + z_k + e_i + r_k) \leq e_i + r_k + T_{ik} + T_{kj}$$

nestes termos, todas as variáveis podem ser interpretadas como rendas de localização, ou mais precisamente: u_i e z_k são custos marginais e também medida das rendas de localização nas zonas de produção e centros de armazenagem. Do mesmo modo v_j é interpretada como renda de localização auferida nos centros de consumo ¹⁴.

O uso de um exemplo ilustrará a interpretação das variáveis do dual como rendas de localização. Suponha-se a existência de dois centros de armazenagem $k = 1, 2$ e um mercado j . Admite-se também, sem perda de generalidade que:

$$T_{i1} > T_{i2} \quad T_{1j} > T_{2j}$$

os preços obtidos no mercado pelo produto oriundo de cada centro de armazenagem são:

$$P_1 = e_i + r_1 + T_{i1} + T_{1j}$$

$$P_2 = e_i + r_2 + T_{i2} + T_{2j}$$

suponha-se também que $r_1 = r_2$, de modo que:

$$P_1 > P_2$$

Se os dois armazéns são necessários para satisfazer a demanda em j , isto é $x_{i1j} > 0$ e $x_{i2j} > 0$, então o preço final no mercado será igual a P_1 , onde:

$$P_1 = e_i + r_1 + T_{i1} + T_{1j} = e_i + r_1 + v_j$$

donde conclui-se que:

$$v_j = T_{i1} + T_{1j}$$

¹⁴ Renda é o pagamento a um fator de produção de oferta inelástica (renda de escassez). Quanto à terra, o pagamento da renda pode ocorrer também em razão de diferenças de fertilidade (renda diferencial) ou ainda como resultado de vantagens locacionais, como por exemplo, a proximidade do centro consumidor (renda de localização); ver STEVENS, B H Linear Programming and Location Rent, *Journal of Regional Science* (Vol 3, n.º 2, 1961)

e, se os custos unitários de produção e armazenagem são iguais para todos os centros o problema se transforma num modelo de transporte com a restrição dual escrita na forma:

$$v_j - u_i - z_k \leq T_{ik} + T_{kj}$$

como $x_{ij} > 0$, obtém-se, pela substituição do valor acima para v_j na restrição do dual a seguinte expressão:

$$(u_i + z_1) = v_j - (T_{i1} + T_{1j}) = 0$$

ou seja: o somatório das rendas de localização em i e $k = 1$ é igual a zero. Levando-se em conta que o custo do transporte é uma medida de distância econômica, vê-se que o centro de armazenagem com maior distância em relação ao mercado auferirá uma renda de localização igual a zero. Por seu turno, quando $k = 2$, onde

$x_{ij} > 0$ obtém-se

$$(u_i + z_k) = v_j - (T_{i2} + T_{2j})$$

substituindo-se novamente o valor de v_j :

$$(u_i + z_2) = (T_{i1} - T_{i2}) + (T_{1j} - T_{2j})$$

isto é:

$$u_i = T_{i1} - T_{i2}$$

e

$$z_2 = T_{1j} - T_{2j}$$

ou seja, u_i e z_2 representam as rendas de localização da zona de produção i e do centro de mercado $k = 2$, medidas em diferenciais (ou vantagens) de custo de transporte. É óbvio que, se a zona i de produção e os centros de armazenagem estiverem todos localizados no mercado j tem-se que:

$$(u_i + z_k) = v_j - T_{jj} = v_j - 0 = v_j$$

o que indica ser v_j a renda de localização no centro de consumo j .

É conveniente que, antes de passar ao caso das culturas temporárias, sejam feitas algumas observações sobre a interpretação da função objetivo do dual e sobre a atuação dos agentes econômicos envolvidos na otimização da produção e distribuição de um produto agrícola. Até este ponto foram implicitamente considerados três agentes: o produtor (agricultor), o consumidor e o agente responsável pela atividade de armazenagem. Suponha-se agora a existência de um quarto tipo de agente denominado "intermediário comercial" cuja função é

comprar o produto na zona produtora, pagar pela sua estocagem e vendê-lo no mercado. O lucro deste intermediário no percurso i, k, j , será pois:

$$l_{ikj} = v_j - z_k - u_i - c_{ikj}$$

ora, o lucro total de intermediação (I) é igual a:

$$I = \sum_{i=1}^m \sum_{k=1}^p \sum_{j=1}^n l_{ikj} x_{ikj} = \sum_{i=1}^m \sum_{k=1}^p \sum_{j=1}^n (v_j - z_k - u_i - c_{ikj}) x_{ikj}$$

ou, desdobrando os somatórios:

$$I = D(v, z, u) - F(X)$$

o que, na solução ótima resulta:

$$I^* = D(v^*, z^*, u^*) - F(X^*) = 0$$

ou seja, o lucro de intermediação é igual a *zero*. Conclui-se que a função objetivo do dual é maximizada de modo a conduzir os intermediários a uma situação ótima onde o lucro é nulo.

Feitas as considerações acima, passa-se agora ao dual do modelo para culturas temporárias. Começa-se por multiplicar as restrições do primal sobre a capacidade de produção e armazenagem por -1 . Feito isto, o dual pode ser escrito como:

$$\text{Max } D(v, z, y, w) = \sum_{j=1}^m d_j v_j + \sum_{k=1}^p -a_k z_k + \sum_{i=1}^n -L_i Y_i - Mw$$

sujeita às restrições:

$$v_j - z_k - g_i y_i - h_i w \leq c_{ikj} \quad i = 1 \dots m, k = 1 \dots p, j = 1 \dots n$$

as variáveis v_j e z_k são interpretadas exatamente como no caso anterior. Para as demais tem-se na solução ótima que:

$$\frac{\partial F(X^*)}{\partial L_i} = y_i^* \quad \frac{\partial F(X^*)}{\partial M} = w^*$$

o que significa dizer que y_i^* representa a renda de escassez da terra na região produtora i , enquanto w^* é a taxa marginal de salário para o conjunto de regiões. É conveniente ressaltar que o valor w^* é um indicador de equilíbrio pois, se numa região i qualquer verifica-se $w_i > w^*$, então este diferencial dá origem a um fluxo migratório na direção de i , pressionando para baixo o salário na região até o ponto em que $w_i = w^*$ para todo i . O mecanismo inverso ocorrerá quando $w_i < w^*$. Conclui-se que a esse nível de salário (w^*), o processo de alocação de mão-de-obra entre as regiões produtoras entra em equilíbrio.

Observa-se ainda que nas restrições do dual as variáveis y_i e w são multiplicadas pelos respectivos coeficientes técnicos. Na verdade, o que a restrição do dual faz é imputar o custo unitário c_{ikj} de volta aos fatores ou recursos envolvidos no processo de produção e distribuição. Além disto, c_{ikj} é expresso em termos de unidades do produto e o mesmo deve ocorrer com as variáveis duais. Sendo y_i expressa em renda paga (cruzeiros) por hectare e w em cruzeiros por homem/dia, ambos devem ser multiplicados pelos respectivos coeficientes técnicos a fim de expressá-los em unidades do produto.

5. APRECIACÃO FINAL

O estudo sobre a aplicação teórica dos modelos multidimensionais à produção e distribuição de produtos agrícolas mostrou o grande potencial dessa classe de modelos, não apenas para melhor compreensão desses processos, mas ainda para possíveis aplicações a problemas de política econômica. A flexibilidade dos modelos multidimensionais permite a representação de problemas práticos complexos, embora isto normalmente acarrete altos custos para sua solução. Outro aspecto não citado neste estudo, é o que se refere à possibilidade de utilizar métodos heurísticos para a análise e solução dos problemas. Em especial nas culturas temporárias a matriz de parâmetros pode ser particionada de maneira a facilitar o uso dos algoritmos para solução.

Através da ótica dual o modelo fornece indicadores da escassez relativa dos recursos e fatores empregados no processo de produção e distribuição, os quais podem ser muito úteis como instrumentos de política econômica. Para um esquema de decisões descentralizadas, as variações nos preços duais permitem simular o impacto de medidas de política sobre o comportamento do setor agrícola¹⁵. Da mesma forma, pode-se aplicar testes de sensibilidade aos parâmetros do modelo, tais como os coeficientes técnicos e de custo. Dentre as principais conclusões do modelo, inclui-se a de que, numa situação de concorrência perfeita, os duais ou rendas diferem entre as zonas de produção. Em outras palavras isto quer dizer que, para aumentar a eficiência do sistema, qualquer política de preços mínimos deve adotar preços diferentes para as zonas produtoras, tomando por base as relações indicadas pelos valores duais.

15 Ver BAUMOL, W & FABIAN, T Decomposition, Pricing for Decentralization and External Economies *Management Science*, Set 1964

ESTRUTURA DO PARQUE CAFEEIRO PARANAENSE (Safrá 1971/72)

Orlando Sá Leite

Estatístico

SUMÁRIO

- 1 *Introdução*
- 2 *Conceitos Básicos Gráfico Ideal*
3. *Estado do Paraná*
4. *Outras Considerações*
5. *Produtividade × Produção*
- 6 *Aplicações*

1. INTRODUÇÃO

As pesquisas sobre Previsão de Safras cafeeiras que o IBC vem aplicando desde 1961, através de modernas técnicas estatísticas de investigação, têm possibilitado o conhecimento dinâmico da realidade cafeeira.

Como ponto de partida, essas pesquisas tinham por finalidade, além das previsões e estimativas das colheitas, responder a perguntas, como por exemplo:

- a) Quantos cafeeiros existem no Brasil? Qual a sua distribuição segundo:
- a.1 — Localização
 - a.2 — Variedade
 - a.3 — Idade
 - a.4 — Produtividade
- b) Qual a distribuição das áreas segundo sua utilização?
- c) Quais os preços pagos aos lavradores?
- d) Qual a quantidade de adubo empregada?
- e) Qual é o estoque de café nas tulhas dos estabelecimentos?
- etc.

A remodelação crescente dos planos de amostragem, a adoção de levantamentos aerofotogramétricos (de forma pioneira no País), a padronização de questionários, o emprego de computadores eletrônicos e a especialização constante de técnicos, permitiu formar, dentro de níveis de precisão adequados, um expressivo elenco de informações estatísticas sobre o Setor da Produção Cafeeira.

Com a arrancada para o plano de racionalização de cafeicultura, entende-se que um papel bem mais importante e complexo estará reservado às pesquisas de safra, na medida em que sejam estabelecidos parâmetros que permitam identificar, medir e acompanhar os resultados dos programas em desenvolvimento.

O trabalho ora apresentado não se prende à análise dos métodos de Amostragem empregados para a obtenção das estimativas, os quais têm sido objeto de outros artigos.

O que se pretende, tão somente, é mostrar o quanto podem ser úteis os levantamentos por Amostragem, desde que os informes pesquisados sejam convenientemente tratados.

O presente estudo, portanto, embora aplicado à cultura cafeeira, pode ser estendido a outros produtos agrícolas, possibilitando uma radiografia da estrutura da produção nos seus mais importantes aspectos.

2. CONCEITOS BÁSICOS GRÁFICO IDEAL

A estrutura ideal de uma população cafeeira, seria aquela em que uma determinada produção fosse proveniente de poucos pés, dotados de alta produtividade.

Portanto, as curvas de distribuição da produção, da população cafeeira e dos estabelecimentos cafeeiros, deveriam apresentar maior concentração nas classes de alta produtividade.

Tomando-se como exemplo a distribuição de uma determinada população cafeeira, segundo classes de produtividade, a qual apresenta-se aproximadamente normal, é possível identificar, pelo aspecto gráfico, algumas situações a seguir representadas:

ESTRUTURA TEÓRICA DE UMA POPULAÇÃO CAFEIEIRA

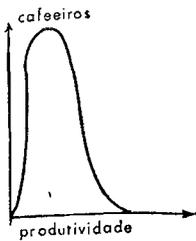


Gráfico 1

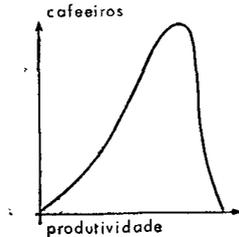


Gráfico 2

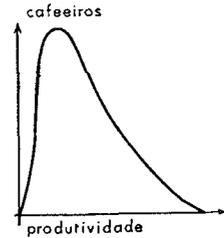


Gráfico 3

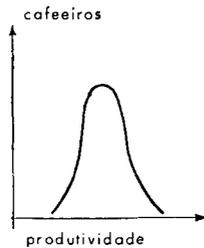


Gráfico 4

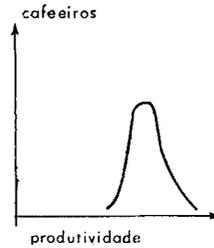


Gráfico 5

A seqüência das imagens gráficas, mostra estruturas cafeeiras que podem ser facilmente classificadas, segundo as variações de produtividade e concentração, como por exemplo:

- GRÁFICO 1 — População simétrica, com alta concentração nas classes de baixa produtividade;
- GRÁFICO 2 — População assimétrica, com alta concentração nas classes de maior produtividade;
- GRÁFICO 3 — População assimétrica, com maior concentração nas classes de menor produtividade;
- GRÁFICO 4 — População normalmente distribuída, com razoável grau de concentração;
- GRÁFICO 5 — População normalmente distribuída, com grande concentração nas classes de alta produtividade.

Os três últimos gráficos representam a evolução de uma população cafeeira a caminho da racionalização, que é conseqüência do “aumento da produtividade e da concentração”.

Seguindo o mesmo raciocínio, por abstração, chegaríamos a uma situação em que toda a população produzisse com o mesmo rendimento (dispersão nula).

A comparação da distribuição dos cafeeiros, com a distribuição da produção e das propriedades, deve ser feita em termos relativos, em virtude da heterogeneidade de unidades.

É importante observar, que nesse caso, os gráficos de cafeeiros e estabelecimentos devem se apresentar com formas semelhantes.

Quando as distribuições são assimétricas, a comparação dos gráficos da população cafeeira e produção são complementares pois "para uma determinada produção, o número de cafeeiros é inversamente proporcional à produtividade".

Isto poderá ser facilmente observado através dos gráficos apresentados no final sobre a estrutura cafeeira do Estado do Paraná.

3. ESTADO DO PARANÁ

"Muitas propriedades, com muitos cafeeiros, produzindo pouco".

"Poucas propriedades, com poucos cafeeiros, produzindo muito".

Afirmações como essas, serão extraídas dos dados e quadros apresentados, que representam as condições existentes durante a safra 1971/72.

3.1 — O QUADRO 1, engloba a distribuição da população cafeeira, da produção e do número dos estabelecimentos cafeeicultores, segundo classes de produtividade em valores simples, acumulados, absolutos e relativos.

3.2 — O GRÁFICO 7, que apresenta as três variáveis em estudo, inter-relacionadas e distribuídas ao longo das classes de produtividade em valores percentuais e acumulados, permite observar:

- a) 71% da população cafeeira está abaixo do nível de 72 sacas coco por 1.000 pés;
- b) essa população é responsável por apenas 49% da produção, para a qual concorrem 81% das propriedades cafeeiras;
- c) conseqüentemente, 51% da produção do Estado é proveniente de apenas 28% da população cafeeira, que se distribui por 19% das propriedades dedicadas a cafeeiculturas.

3.3 — OS GRÁFICOS 7, 8 e 10 apresentam as distribuições de frequência das variáveis consideradas, tornando possível a identificação de outros aspectos, como por exemplo:

3.3.1 — GRÁFICO n.º 8 — População Cafeeira — 54% dos pés encontram-se no intervalo: 36 a 90 sacas coco por 1.000 pés;

3.3.2 — GRÁFICO n.º 9 — relativamente à produção, nota-se que 76% da mesma é proveniente de cafeeiros com produtividade na faixa 36 a 108 sacas coco por 1.000 pés;

3.3.3 — GRÁFICO n.º 10 — a distribuição das propriedades demonstra que a maioria dos estabelecimentos apresenta baixa produtividade: 71% das propriedades encontram-se nas classes inferiores a 72 sacas coco ou, aproximadamente, 24 sacas beneficiadas.

4. OUTRAS CONSIDERAÇÕES

Considerando que o tamanho das propriedades é diretamente proporcional à capacidade produtiva, pode-se concluir que:

4.1 — 50% da produção do Estado encontra-se comandada por apenas 19% dos estabelecimentos cafeicultores.

4.1.1 — Nesse grupo, a média de pés (covas) por propriedade, é de 14 mil.

4.2 — Para o Estado, a média de pés (covas) por propriedade é de 9 mil.

4.3 — Os outros 50% da produção do Estado são provenientes de 80% dos estabelecimentos cafeicultores.

4.3.1 — Nesse grupo a média de pés (covas) por estabelecimentos é inferior à média global do Estado (8 mil).

4.4 — A classe de produtividade que mais colabora para a produção é: 72 a 90 sacas coco/1.000 pés.

4.5 — A produtividade média da colheita em 1971 foi de 72 sacas coco/1.000 pés.

4.6 — A classe de produtividade que concentra maior número de estabelecimentos é: 36 a 54 sacas.

4.7 — A produtividade média das propriedades foi de 46 sacas coco por 1.000 pés.

Através da reunião de algumas classes, que delimitam os níveis de produtividade, pode-se estabelecer uma comparação sintomática da estrutura da cafeicultura, como por exemplo:

Intervalos	População M	Produção	N.º de Propriedades
1 a 36	32,6%	11,8%	43,9%
36 a 90	52,8%	28,4%	45,7%
90 a 180	14,6%	29,8%	10,4%

ESTRUTURA DA CAFEICULTURA
PARANAENSE SEGUNDO NÍVEIS
DE PRODUTIVIDADE

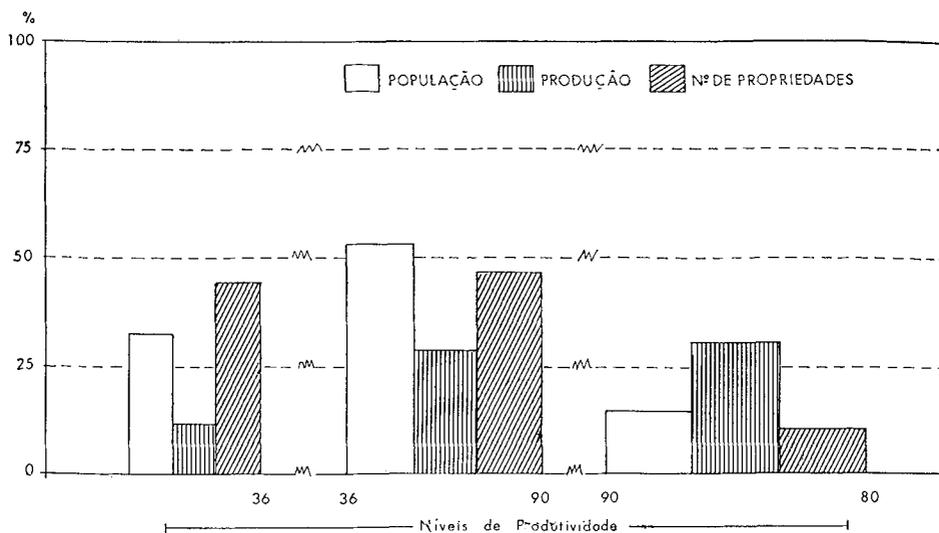


Gráfico 6

A observação do Gráfico 6, demonstra a afirmativa feita anteriormente, ou seja “muitos cafeeiros e propriedades com baixa produção e poucos cafeeiros e propriedades com alta produção”.

5. PRODUTIVIDADE X PRODUÇÃO

Teoricamente, a elevação do nível de produtividade da população cafeeira paranaense aumentaria a oferta consideravelmente.

Constatou-se na safra de 1971, que 231 milhões de pés tiveram produtividade de 19,7 sacas, contra 54 sacas que foi a média do Estado. Essa população cafeeira foi responsável pela produção de 4,5 milhões de sacas (11,8% da produção da safra).

Admitindo-se que tal conjunto de cafeeiros atingisse a média de produtividade do Estado, a produção correspondente seria acrescida de 8 milhões de sacas.

6. APLICAÇÕES

O tipo de análise apresentado, poderá, ainda fornecer outras importantes informações como por exemplo:

- Estabelecer comparação entre os diferentes Estados produtores em todas as safras;
- Acompanhar o nível de racionalização da população cafeeira, através de confronto dos gráficos apresentados;
- Orientar a adoção de programas específicos destinados ao aumento do nível de produtividade das lavouras;
- Estimar, com grande precisão, os efeitos de geadas, através da identificação das classes produtivas atingidas;
- Subsidiar os estudos sobre Custo de Produção, relativamente à fixação de classes de produtividade mais representativas, etc.

QUADRO N.º 1

DISTRIBUIÇÃO DA POPULAÇÃO CAFFEEIRA, PRODUÇÕES E PROPRIEDADES, SEGUNDO NÍVEIS DE PRODUTIVIDADE NO ESTADO DO PARANÁ

CLASSES DE PRODUTIVIDADE (*)	CAFFEEIROS ADULTOS COM PRODUÇÃO		PRODUÇÃO SACAS COCO		PROPRIEDADE EXISTENTE	
	Quantidade	%	Quantidade	%	Quantidade	%
1 — 9	33 080 248	4,6	167 889	0,4	7 407	9,2
9 — 18	75 821 578	10,7	1 006 029	2,6	10 815	13,5
SUB TOTAL	108 901 826	15,3	1 173 918	3,0	18 222	22,7
18 — 27	57 087 366	8,0	1 347 533	3,5	7 863	9,9
SUB TOTAL	165 989 192	23,3	2 521 451	6,5	26 085	32,6
27 — 36	65 462 934	9,3	2 020 793	5,3	9 083	11,3
SUB TOTAL	231 452 126	32,6	4 542 244	11,8	35 168	43,9
36 — 54	163 017 026	22,9	7 287 211	19,0	18 342	22,9
SUB TOTAL	394 469 152	55,5	11 829 455	30,8	53 510	66,8
54 — 72	110 724 394	15,6	6 910 719	17,9	11 553	14,5
SUB TOTAL	505 193 546	71,1	18 740 174	48,7	65 063	81,3
72 — 90	101 308 386	14,3	8 251 136	21,5	6 689	8,3
SUB TOTAL	606 501 932	85,4	26 991 310	70,2	71 752	89,6
90 — 108	68 604 413	9,7	6 802 467	17,7	5 664	7,1
SUB TOTAL	675 106 345	95,1	33 793 777	87,9	77 416	96,7
108 — 144	23 833 386	3,4	3 016 908	7,9	1 358	1,7
SUB TOTAL	698 939 731	98,5	36 819 685	95,8	78 774	98,4
144 — 180	10 637 412	1,5	1 593 261	4,2	1 225	1,6
TOTAL	709 577 143	100,0	38 403 946	100,0	79 999	100,0

(*) Número de sacas coco por mil pés.

FONTE: IBC/DEC — Divisão de Estatística.

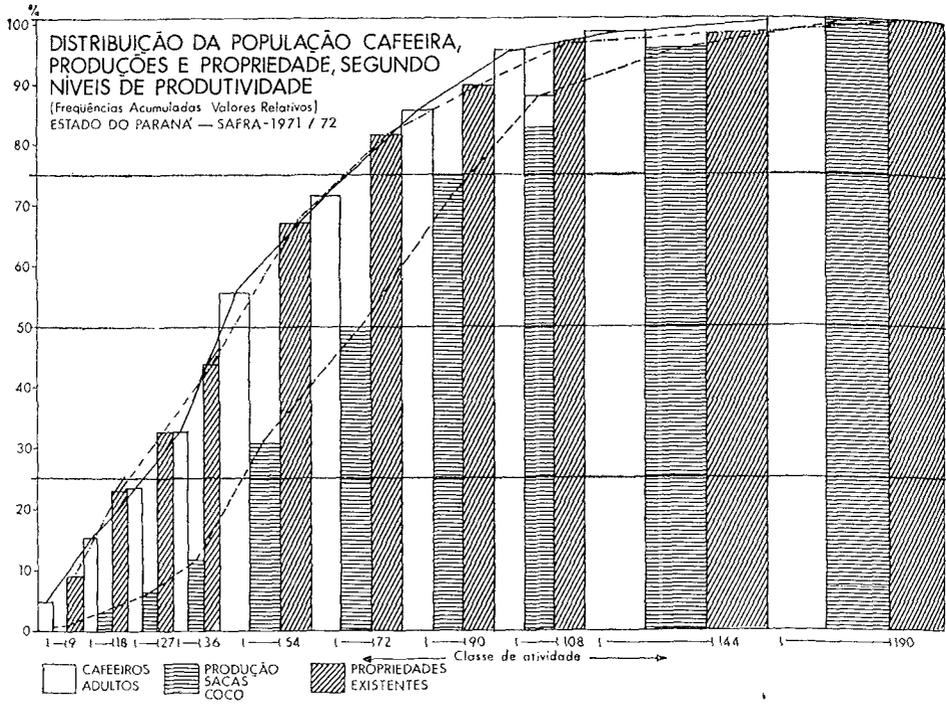


Gráfico 7

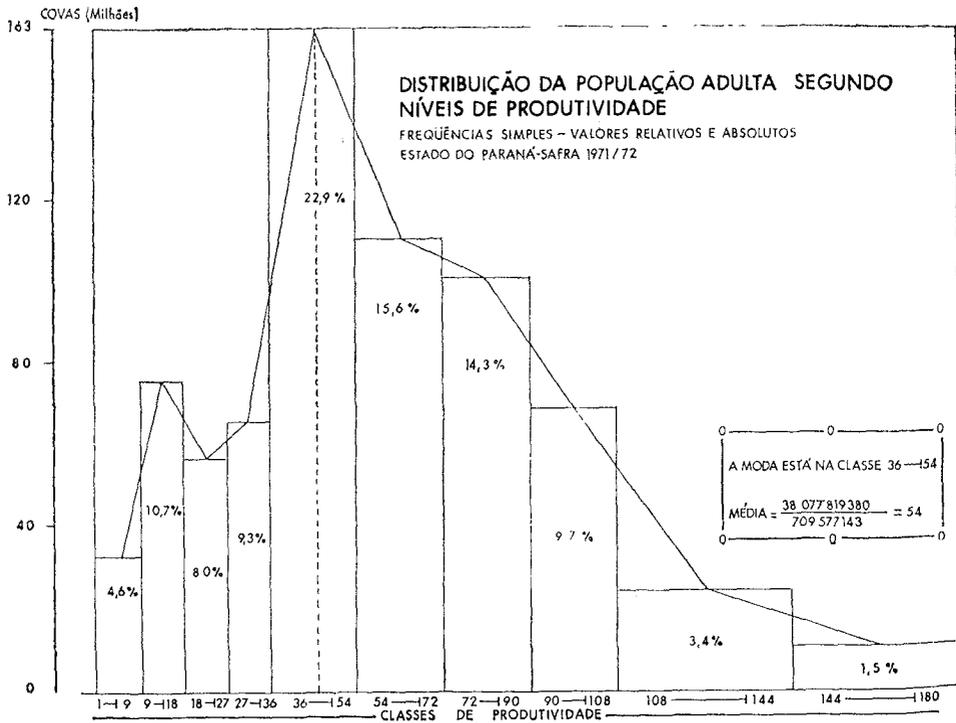


Gráfico 8

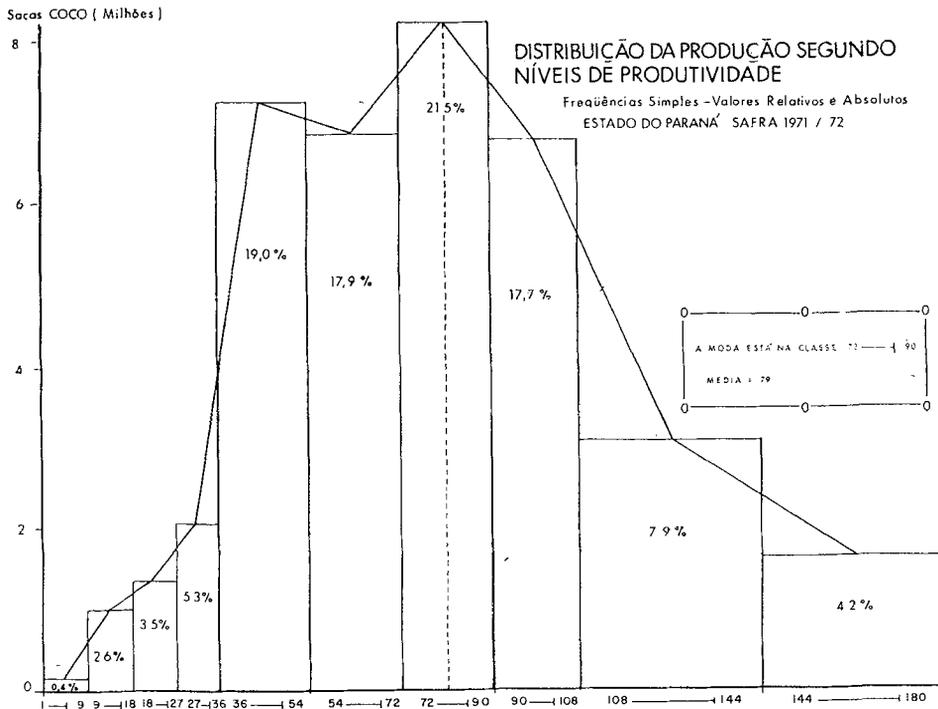


Gráfico 9

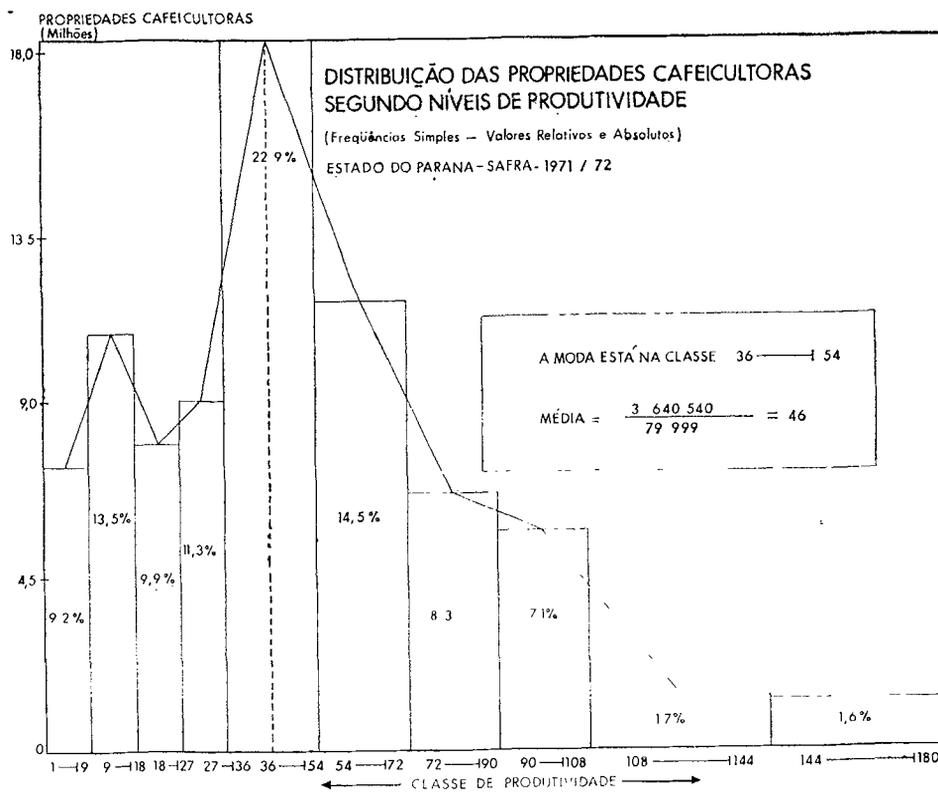


Gráfico 10

ORGANIZAÇÃO E MÉTODOS DO SISTEMA DUAL DE REGISTROS NA ÍNDIA*

Joan Lingner
e
H. Bradley Wells

SUMÁRIO

- I *Introdução*
- II *Índia: Síntese*
- III *Origens e Desenvolvimento do Sistema Dual de Registro na Índia*
- IV *Organização Administrativa: Retrospecto*
- V. *Treinamento de Pessoal de Campo*
- VI *O Plano da Amostra*
- VII *Mapeamento e Outros Processos Iniciais*
- VIII *O Processo de Registro Contínuo*
- IX. *A Entrevista do Levantamento*
- X. *Circulação de Documentos e Crítica Inicial*
- XI *Comparação e Variação de Campo*
- XII. *Tabulação e Análise*
- XIII *Relacionamento do Sistema Dual de Registro Indiano com Outras Atividades de Coleta de Dados*
- XIV *Estudos Analíticos e de Avaliação*
- XV *Fontes de referências*

* Transcrição de "Organization and Methods of the Dual Record System in India", do Laboratories for Population Statistics, Scientific Report Séries n° 9, October 1973. Tradução de Maria Luiza Maier Pollnov, com autorização do International Program of Laboratories for Population Statistics (POPLAB).

I. INTRODUÇÃO

O estudo intensivo das modificações de população, em decorrência do desenvolvimento econômico e social, forçou os demógrafos a reconsiderarem a adequação dos métodos tradicionais de mensuração dessas mudanças. Mesmo altamente desenvolvidos, censos e sistemas de registro demográfico não proporcionam, no devido tempo, a profundidade de conteúdo necessária para preencher os dados básicos ao planejamento e programas de avaliação. A situação é, naturalmente, muito pior nos países em que os sistemas censitários e de registro vitais são inexistentes ou muito deficientes. Está havendo progresso para uma cobertura melhor, tanto pelo censo, como pelos sistemas de registros demográficos, mas, na maioria dos países, serão necessários décadas para se atingir uma cobertura total (completa).

Foi proposto e testado, como alternativa dos métodos tradicionais, métodos alternativos para coleta de dados demográficos, principalmente sobre nascimentos e óbitos. Incluem eles: sistemas duais de registro, levantamentos gerais retrospectivos, simples e múltiplos e tentativas de fortalecer os sistemas de registro civil já existentes, através de amostras.

No sistema dual de registro, os dados sobre acontecimentos demográficos e as características daí resultantes são coletados simultaneamente através de: 1) um processo contínuo de observação e registro e 2) uma série de levantamentos retrospectivos. Os levantamentos dão também informações sobre o tamanho e características da população base. Já que cada evento é lançado duas vezes, sendo uma pelo processo de registro contínuo e a outra pelo levantamento retrospectivo, os eventos individuais apanhados por um dos componentes do sistema podem ser comparados com os encontrados pelo outro. A verificação comparativa cruzada pode, às vezes, ser uma característica dos levantamentos gerais múltiplos ou das amostras de registro civil, mas está sempre presente nos sistemas duais de registros. Conseqüentemente, os sistemas duais de registro, neste sentido, são auto-controladores.

Experimentos de âmbito nacional, aplicando a metodologia do registro dual, têm sido, ou estão sendo, efetuados na Índia, Tailândia, Libéria, Paquistão e Turquia (Libéria, 1969; Instituto Paquistanês de Desenvolvimento Econômico, 1968; Rumford, 1970; Rumford et al., 1968; Lauriat, 1967). Relatórios anteriores nestas Séries Científicas das publicações da POPLAB descreveram o desenvolvimento, a organização e os métodos experimentais realizados na Libéria, Paquistão e Turquia (Albernathy, 1972; Chanlett, 1971; Lingner, 1972). O presente relatório focaliza o sistema dual indiano de registro, acompanhando sua evolução por quase uma década. O sistema indiano é chamado de *Sistema de Registro Por Amostras*, embora seja independente do registro civil.

Foram realizados estudos comparativos destes testes, em larga escala, bem como abordagens relacionadas com mensuração demográfica (Lauriat, 1967; Seltzer, 1969; Wells, 1971; Abernathy e Lunde, 1972); entretanto, a natureza essencialmente semelhante destas abordagens tornou-as um tanto vagas pela diversidade de terminologia usada na descrição dos vários processos. Conseqüentemente, neste relatório permitimo-nos certas liberdades quanto aos termos usados no sistema dual indiano de registro. Por exemplo, usamos o termo *sistema dual de registro*, em vez de *sistema de registro por amostra*; a pessoa que tem a tarefa de observar continuamente e fazer o registro é chamada de *registrador neste relatório* (no sistema indiano é o *agente censitário*); a pessoa que dirige as pesquisas é o *entrevistador de pesquisa* (*supervisor* no sistema indiano). Isto foi feito para se manter coerência com os relatórios publicados anteriormente nesta série.

Este estudo baseia-se principalmente nos relatórios publicados no Sistema de Registro por Amostra da Índia. Agradecemos ao Sr. A. Chandra Sekhar, Registrador Geral, Sr. V. K. Ramabhadran, Registrador Geral Substituto, Sr. G. A. Kulkarni, Registrador Geral Assistente, Sr. R. D. Agrawal e Sr. V. S. Swamy, Agentes Sênior de Pesquisa, e a outros funcionários da equipe da Unidade de Registro por Amostra, por sua assistência em proporcionar-nos informações e comentários acerca de um plano anterior. Quaisquer equívocos ou afirmações errôneas são, naturalmente, da inteira responsabilidade dos autores.

II. ÍNDIA: SÍNTESE

Ocupando a maior parte do subcontinente sul-asiático, a Índia caracteriza-se por sua diversidade climática, topográfica e social, e pela continuidade de sua longa história cultural. Geograficamente, o país ocupa uma área de 1,2 milhão de milhas quadradas (3 milhões de km²) e tem o formato aproximado de um losango, com ângulos em cada um dos pontos cardiais. A noroeste, a Índia, limita-se com o Paquistão, no nordeste com a China, Tibet, Nepal e os Estados semi-independentes de Sikim e Butan, e a fronteira do sudeste é formada por Burma, Bangladesh e a Baía de Bengala. O Mar da Arábia delimita a Índia pelo sudoeste.

Pode-se identificar três amplas regiões topográficas. a região montanhosa setentrional da cadeia do Himalaia, a área das planícies, que se estende desde as terras desérticas de Rajastan, a noroeste, em direção sudoeste ao longo do Ganges e seus tributários até a Bengala Ocidental, e o platô de Decan, que predomina na península sulina. Embora o clima varie de tropical a temperado, pode-se distinguir 3 estações para a maior parte do país: uma estação quente seca, de março a junho, uma estação chuvosa, de junho a novembro, e um estação tem-

perada para os meses restantes. O nível pluviométrico varia muito, de aproximadamente 400 polegadas anuais, nas montanhas de Assam, para menos de 5 polegadas, nas regiões desérticas do sudoeste.

O Censo da Índia de 1971 revelou uma população total de 548 milhões de habitantes, colocando a Índia em 2.º lugar, logo após a China, em tamanho de população. Devido aos elevados coeficientes de mortalidade, o crescimento vegetativo foi relativamente lento até 1920. Nos últimos 50 anos, entretanto, a mortalidade decresceu de uma taxa estimada de mais de 40 por mil para um nível aproximado de 18, em 1969. O coeficiente de natalidade, entretanto, aparentemente só caiu levemente de um nível aproximado de 45 por mil, antes de 1940, para o nível atual de cerca de 38 (Índia, Vital Statistics Division, 1972 b). Como consequência do declínio mais rápido na mortalidade do que na fecundidade, a população da Índia está aumentando presentemente a uma taxa de cerca de dois por cento ao ano.

A economia da Índia é em grande parte agrícola, com quase três quartos da população economicamente ativa, em 1961, ocupada neste setor. No setor não-agrícola da economia, a indústria têxtil predomina tradicionalmente, mas outras indústrias, abrangendo desde a extração de matérias-primas até a manufatura de produtos finais, vem se desenvolvendo rapidamente.

A história política e social do subcontinente indiano deve ser considerada em termos de milênios. Antes da era moderna, os períodos de unificação alternavam-se com períodos de fragmentação em pequenos Estados. Durante o período de transição, no século XVIII, quando o Império Mongol entrou em decadência, foi introduzido o poder colonial inglês, que durou até 1947, quando foi conseguida a independência. O subcontinente foi então, dividido nas duas nações: Índia e Paquistão (este último compreendendo agora o Paquistão e Bangladesh).

A Índia é uma democracia parlamentar, com um governo central relativamente forte que reúne atualmente (1973) 21 Estados, cada qual com seu próprio parlamento de duas câmaras, e oito Territórios Federais. Os Estados e Territórios são divididos em cerca de 350 distritos administrativos; estes, por sua vez, subdividem-se em quadras (blocos), contendo cada uma 50 a 100 mil pessoas.

Atualmente há 14 idiomas oficialmente reconhecidos e cerca de 850 dialetos, representando duas famílias lingüísticas principais; as línguas indo-européias predominam no norte e as dravidianas no sul. O hinduísmo é a religião de 83 por cento da população; outros grupos religiosos são: os muçulmanos (11%), cristãos (3%), sikhs (2%), budistas e jains (1%) (Índia, Office of the Registrar General, 1972 b).

Em vista da economia essencialmente agrícola da Índia, a maioria da população vive em cerca de meio milhão de aldeias rurais, muitas delas relativamente isoladas e economicamente independentes. Somente cerca de 20% da população vive em áreas urbanas.

III. ORIGENS E DESENVOLVIMENTO DO SISTEMA DUAL DE REGISTRO NA ÍNDIA

Embora as contagens e estimativas fossem preparadas desde eras remotas, o primeiro censo moderno, abrangendo a maior parte da Índia, realizou-se no período de 1867-72 e é comumente citado como o Censo de 1871. De 1881 até o presente, os censos têm sido realizados em intervalos decenais, mas “até 1951, a Organização do Censo, na Índia, funcionava como a fênix, isto é, a Organização era criada às vésperas dos censos e terminava tão logo as operações censitárias estivessem concluídas” (Srivastava, 1971, pp. 11-12). Em 1958, foi instalada uma organização censitária permanente na Repartição Geral do Registro, no Ministério do Interior (Saxena, 1971).

Os rudimentos de um sistema legal de registro demográfico surgiram em fins do século XIX, com a Lei do Registro de Nascimentos, Óbitos e Casamentos, de 1886. Esta Lei, entretanto, dispunha apenas sobre o registro voluntário de eventos vitais. Diversas leis estaduais e locais tornaram os registros de nascimento e óbito obrigatórios, mas não havia uniformidade no processamento, nem na execução. A legislação nacional, exigindo o registro de nascimentos e óbitos, não pode ser decretada antes de 1969. Embora a Lei do Registro de Nascimentos e Óbitos, de 1969, esteja em vigor, a falta de fortes motivos para o registro voluntário e as dificuldades de comunicação acarretam um registro ainda grandemente deficiente na maioria das áreas e de pouco valor para o fornecimento das estimativas necessárias para coeficientes de natalidade e mortalidade.

Em 1961, uma conferência de profissionais interessados em estatísticas demográficas recomendou uma grande variedade de medidas destinadas a reforçar o sistema de registro demográfico na Índia. Reconhecendo que o aperfeiçoamento no sistema de registro civil em sua totalidade era um processo lento, a conferência insistiu em que se fizesse um plano especial de amostra de registro, capaz de satisfazer às necessidades imediatas de dados sobre nascimentos e óbitos. Esta sugestão foi ainda reforçada pelas recomendações de Linder (1964), que organizou um programa de aperfeiçoamento da contagem de eventos vitais.

A Repartição do Registro Geral da Índia, depois de autorizada pela Comissão de Planejamento e pelo Ministério do Interior a seguir estas recomendações, fez planos detalhados de implementação, tanto para o aperfeiçoamento do sistema de registro demográfico, como para a iniciação do sistema dual de registro. Este último, chamado de Sistema de Registro por Amostras (esquema), é o tema do presente trabalho. Ele é ímpar entre os sistemas de registro dual, não só pelo tamanho da amostra, aproximadamente 3,5 milhões, como pela demora de sua operação, que está atingindo uma década.

O sistema dual de registro indiano formulou-se originalmente como um plano conjunto do governo central e dos governos de cada Estado, arcando o governo central com 75% da despesa total nos Estados. Em 1963, cada Estado foi solicitado a iniciar a implementação do sistema dual de registro através de um estudo-piloto, em janeiro de 1964, num número limitado de áreas ou unidades de amostra. A fase-piloto, que geralmente abrangia 20 localidades rurais e 5 a 10 áreas urbanas, deveria ser seguida, depois de um período de um ano, pelo início de um projeto integral, compreendendo habitualmente 150 unidades rurais de amostra e 60 urbanas. Em 1964, tais planos-piloto foram introduzidos em áreas rurais e urbanas de 5 Estados (Bihar, Gujarat, Kerala, Maharashtra e Misore). Mais 2 Estados (Tamil Nadu e Punjab) iniciaram planos-piloto em 1965, e outros aderiram ao programa em anos posteriores.

As repartições encarregadas do sistema dual de registro variavam de Estado para Estado: em alguns, o trabalho era feito pelo Bureau de Economia e Estatística; noutros, o Ministério da Saúde realizava o programa de trabalho. O ritmo da implementação foi lento porque os Estados participantes foram solicitados a contribuir para as despesas nos primeiros anos.

A Repartição do Registro Geral defrontou-se, por isso, com o problema de convencer os funcionários estaduais da necessidade de registro por amostra (Wells e Agrawal, 1967). Em 1969, tornou-se patente que os poucos Estados não participantes do plano provavelmente não o seriam em futuro previsível. A Repartição do Registro Geral decidiu então que os funcionários estaduais do censo empreenderiam o trabalho de campo nos Estados que restavam.

A implementação em áreas urbanas apresentou problemas adicionais, uma vez que os estudos-piloto revelaram que os resultados aí eram menos fidedignos do que os das áreas rurais. Segundo exame mais detalhado, que será feito a seguir, a experimentação com fórmulas alternativas foi efetuada com objetivo de atenuar estas dificuldades. Conforme se vê na Tabela I, as agências censitárias estaduais agora assumiram a responsabilidade sobre uma parcela maior do trabalho de campo na área urbana do que na rural.

Em 1970, a participação no plano foi essencialmente completa para todos os Estados (exceto para Nagaland) e principais territórios da união. Os territórios restantes da união foram incluídos em 1970-71. Para detalhes sobre a data do início de operações em escala total e as agências executantes em cada Estado, ver Tabela I.

IV. ORGANIZAÇÃO ADMINISTRATIVA: RETROSPECTO

Como uma experiência cooperativa entre o governo central e os Estados, o sistema dual de registro indiano requer uma organização bem coordenada e de muitos níveis. Em nível nacional, a responsabilidade

pela coordenação e esquema global cabe à Repartição do Registro Geral, operando através da Unidade de Registro por Amostras da Divisão de Estatísticas Demográficas.

Em 1972, a equipe desta unidade compunha-se de “2 funcionários-sênior de pesquisa, 3 funcionários de pesquisa e outro pessoal técnico, sob a orientação geral do Registrador Geral Substituto (Estatísticas Demográficas)” (Índia, Vital Statistics Division 1972a, p. 8). As responsabilidades principais da Unidade de Registro por Amostra incluem: preparação do plano e seleção da amostra; preparação e revisão de versões, em inglês, dos modelos e questionários, bem como de manuais com instruções para os vários níveis de pessoal; manutenção de um sistema de controle que inclua a identificação de unidades duvidosas ou deficientes; visitas locais frequentes e regulares; recebimento e publicação de relatórios estaduais; tabulação e análise de dados; e preparação de relatórios.

O sistema dual de registro indiano é uma grande e complexa organização. Muito tempo, reflexão, e esforço foram dedicados à identificação e solução de problemas administrativos; ao desenvolvimento e manutenção de padrões uniformes, e à avaliação da eficácia do próprio sistema. Um aspecto importante deste processo é a série de Conferências de Registro por Amostras anuais de representantes dos órgãos estaduais de implementação e das organizações correlatas do governo central. Estas conferências, patrocinadas e coordenadas pela Repartição Geral de Registro, contribuíram para a identificação de problemas técnicos e de obstáculos administrativos e levaram a diversas modificações do sistema. Outro instrumento extremamente valioso no processo de aperfeiçoamento do sistema é a série de estudos de análise e avaliação realizados em nível estadual e nacional. Eles serão examinados mais detalhadamente na Seção XIV deste trabalho. O *Sample Registration Bulletin*, publicado (mensalmente 1967-1970 e depois trimestralmente) pela Repartição Geral de Registro, proporciona coordenação e comunicação e serve como veículo de divulgação dos principais resultados dos estudos de análise e avaliação nos mais breves intervalos de tempo possíveis.

Uma inovação recente no sistema dual de registro indiano é a criação de um grupo de estudos, composto de funcionários da Unidade de Registro por Amostra e de consultores de fora. Este grupo faz visitas locais intensivas a diversas áreas estaduais e locais num esforço de manter e melhorar os padrões de qualidade. O grupo de estudos prepara um relatório e suas recomendações para o aperfeiçoamento da qualidade do trabalho de campo são comunicadas ao órgão de implementação.

Dentre de cada Estado a composição da equipe de trabalho varia de acordo com o órgão responsável pela implementação do plano. Em nível estadual, o órgão de implementação é tipicamente constituído de uma unidade central, que geralmente consiste de um Pesquisador Sênior, em horário integral, ou de um Oficial de Tabulação, 2 Assistentes de Estatística, um Calculador e um Compilador Assistente, tra-

balhando sob a direção geral de um funcionário Sênior (Oficial Estadual de Supervisão) do órgão implementador. Esta equipe central é responsável pela identificação de campo das unidades de amostra selecionadas pela Repartição Geral de Registro; pelo mapeamento e segmentação das unidades maiores; pela tradução para o idioma local, impressão e distribuição de formulários e questionários; e pela seleção, treinamento e controle do pessoal de campo, tanto nas operações de registro, como nas de levantamento. Em geral, as unidades de nível estadual costumam assumir a responsabilidade contínua do recebimento dos formulários preenchidos do campo e da crítica inicial dos dados, comparação, e verificação de campo. O pessoal da esfera estadual também compila e analisa os dados, prepara relatórios estaduais para encaminhamento ao Registro Geral, e realiza diversos estudos de avaliação.

Em nível distrital ou de grupamentos, a equipe existente de cada órgão de implementação foi inicialmente utilizada, em horário parcial, para desempenhar grande variedade de funções de supervisão, inclusive a supervisão imediata do trabalho dos registradores. As equipes distritais ou de grupamentos eram também usadas como entrevistadores no levantamento semi-anual. A experiência adquirida com o plano indicou a necessidade de um controle mais intensivo das atividades de campo; para tanto, foi recrutada uma equipe, em regime de horário integral, para incumbir-se do processo de registro e também do inquérito semi-anual. Atualmente, uma pessoa do padrão de um Calculador, é designada, em regime de tempo integral, para supervisionar os registradores em cerca de 12 unidades de amostra; a mesma pessoa terá a tarefa de fazer as entrevistas para o inquérito semi-anual em 12 outras unidades. Esta separação das atividades de supervisão e entrevista representa um esforço para se manter um maior grau de independência entre o trabalho dos registradores e o dos entrevistadores.

Os registradores formam a coluna mestra do sistema dual de registro indiano. Na maior parte, são eles empregados em regime de tempo parcial, que vivem e trabalham nas unidades de amostra; poucos registradores são empregados, em tempo integral, nas cidades maiores. Em geral foi dada preferência ao recrutamento do professor primário local como registrador, mas foram também consideradas outras pessoas instruídas, desempenhando funções ativas na vida da comunidade, como, por exemplo: agentes de correio, parteiras, assistentes sociais, pessoal da saúde pública e do desenvolvimento. Os registradores recebem baixos honorários por seu trabalho, variando a quantia com o tamanho da unidade de amostra.

O trabalho de campo em cada amostra abrange 5 elementos principais:

- 1) mapeamento detalhado de cada unidade de amostra e inquérito básico para obter a habitual população de direito da unidade;

- 2) observação e registro contínuos dos eventos vitais, que ocorrem dentro da unidade, pelo registrador residente, que localiza os nascimentos e os óbitos através de um sistema de informantes, visitas de casa em casa, ou outros meios, obtendo os detalhes por meio de contato com os domicílios onde ocorreram os nascimentos e óbitos;
- 3) inquéritos semi-anuais independentes, que incluem quesitos retrospectivos sobre nascimentos e óbitos (estes inquéritos também atualizam as informações sobre a habitual população de direito);
- 4) comparação dos eventos verificados através do processo de registro contínuo com os descobertos por meio dos inquéritos semi-anuais;
- 5) verificação em campo de eventos não comparados e parcialmente comparados.

Cada procedimento destes será mais completamente examinado em capítulos subseqüentes deste estudo.

Talvez se deva frisar que o experimento indiano difere num aspecto de outras tentativas para um sistema dual de registro. Reconhecendo que os métodos utilizados no sistema dual de registro indiano não correspondem bem às expectativas do modelo estatístico que serve de base à fórmula Chandra Sekar-Deming (Chandra Sekar e Deming, 1949), não foi feita nenhuma tentativa de ajustar os coeficientes demográficos estimados para eventos omitidos em ambos os componentes do sistema dual de registro (Ramabhadran, 1971). Conforme exame mais completo no Capítulo XI, a aplicação da fórmula Chandra Sekar-Deming aumentaria em muito pouco os coeficientes demográficos estimados.

V. TREINAMENTO DE PESSOAL DE CAMPO

O treinamento conjunto dos entrevistadores de inquéritos e dos registradores é efetuado pela equipe da unidade central do Estado, em vários centros regionais dentro de cada Estado, com o auxílio de funcionários da Repartição Geral de Registro. Os locais para treinamento são geralmente selecionados entre os escritórios distritais da repartição implementadora, visando reduzir ao mínimo o tempo e as despesas de viagem do candidato. Os registradores e entrevistadores de inquéritos recebem manuais de instrução antes do início do treinamento, que devem ser por eles estudados antes de se apresentarem às sessões de treinamento. A duração do programa em cada centro está prevista para 3 dias.

As sessões de treinamento começam com um debate sobre os objetivos do inquérito e a importância dos registradores e entrevistadores para o sucesso do programa. As tarefas do registrador são apresentadas

em detalhe e incluem. métodos para o estudo da estrutura social, usos e costumes em cada unidade e a relação destes fatores para a seleção de informantes; os métodos adequados para completar os diversos formulários utilizados no processo de registro; e o relacionamento entre registradores e supervisores. O papel dos entrevistadores do inquérito é igualmente discutido e eles recebem instruções exatas sobre seus deveres de supervisão, inclusive o de completar os formulários de inspeção.

Esta fase introdutória é seguida de entrevistas práticas, tanto na sala de aula, como no campo. Foi enfatizada a compreensão dos conceitos e definições. Tanto os registradores, como os entrevistadores do inquérito, recebem treinamento no uso de algarismos arábicos para facilitar a tabulação dos dados. Uma sessão final abrange áreas de problemas específicos e inclui um resumo geral da sessão de treinamento. É também feita a previsão para um treinamento suplementar, com outra sessão de treinamento, se o trabalho dos registradores e entrevistas de inquérito demonstrarem deficiência. Sessões de treinamento sumárias para entrevistadores são também realizadas antes de cada inquérito semi-anual.

VI. O PLANO DA AMOSTRA

O objetivo primário do sistema dual de registro na Índia é proporcionar estimativas fidedignas de coeficientes de nascimentos e óbitos não só para a nação como um todo, mas também para cada Estado e para as áreas urbanas e rurais em separado. Por isso, o sistema fornece bases para o estudo de diferenças nos coeficientes de fecundidade e mortalidade estaduais, regionais e urbanos-rurais.

O plano da amostra e a seleção são executados em nível nacional pela Repartição de Registro Geral; a verificação de campo e o mapeamento das unidades de amostra são feitos pelas entidades implementadoras de cada Estado.

O universo rural inclui toda a população rural, excetuando-se algumas partes de Jamu e Caxemira e Assam, excluídas por causa de dificuldades operacionais. Instruções detalhadas para estruturação dos fundamentos da amostragem urbana e rural são dadas no *Headquarters' Manual* (Índia, Vital Statistics Division, n. d. a.).

Na maioria dos Estados, o plano de amostra rural completa consiste numa amostra aleatória, estratificada, unifásica, de 150 vilas ou setores de vilas; em alguns dos Estados menores e nos Territórios da união, o tamanho da amostra oscila entre 40 e 100 unidades (ver Tabela I). O *1961 Census Village Directory* (Anuário do Censo das Vilas, de 1961) serviu como fundamento da amostragem. A estratificação foi realizada primeiramente através da subdivisão da área rural total den-

tro de cada Estado em quatro divisões naturais, i.é., sub-regiões geográficas. Dentro destas subdivisões, as vilas foram classificadas, segundo o número de habitantes, em 4 categorias:

menos de 500
500 — 999
1.000 — 1.999
2.000 e mais

As 150 unidades rurais de amostra foram destinadas a cada estrato em proporção à população de 1961. Para as primeiras três categorias, as vilas de amostra foram selecionadas aleatoriamente dentro de cada divisão geográfica.

Para assegurar unidades de amostra de tamanho viável, adotou-se um esquema de amostragem um pouco diferente para vilas maiores. No estrato de vilas de 2.000 e mais habitantes, cada vila compreendia duas ou mais unidades de amostra. O número de unidades fixado para cada vila foi determinado pelo menor número inteiro encontrado que, na divisão da população das vilas, de 1961, apresentasse unidades de menos de 2.000 habitantes. Por exemplo, às vilas de 2.000 — 3.999 habitantes, em 1961, foram atribuídas duas unidades, e vilas de 4.000 — 5.999 habitantes teriam três unidades. O número indicado de unidades de amostra foi então selecionado aleatoriamente. Assim, o mapeamento e zoneamento efetivo para o número de unidades indicado só se faziam necessários para aquelas vilas que tinham uma unidade escolhida na amostra. A identificação da unidade específica a ser incluída na amostra, dentre os setores mapeados, foi feita no escritório central.

Nos Estados que abrangem 150 unidades rurais, os erros-padrão relativos variam entre 1,6 a 2,8% para os nascimentos e de 3,0 a 4,6% para os óbitos (Índia, Vital Statistics Division, 1972a).

O universo urbano é o total da população urbana em cada Estado. Um plano em duas etapas foi empregado na seleção da amostra urbana. Grandes cidades e cidades foram primeiramente classificadas em quatro estratos, de acordo com o tamanho da população em 1961:

100.000 e mais
50.000 — 99.999
20.000 — 49.999
menos de 20.000

As unidades de amostra foram distribuídas nestes estratos em proporção à população; entretanto, pelo menos duas unidades de amostragem foram determinadas para cada cidade no primeiro estrato e, em alguns Estados, no segundo também. Com exceção dos dois primeiros estratos, as grandes cidades e as cidades serviram como unida-

des de amostragem de primeira etapa e os blocos (quadras) censitários dentro das cidades formaram a segunda etapa. No conjunto 60-100 blocos foram selecionados em cada Estado (ver Tabela I).

Nas fases-piloto do projeto, foram efetuados vários experimentos com o tamanho das unidades de amostra urbana. Assim, os cinco Estados que iniciaram a implementação do projeto em 1964 usaram unidades de 10.000 habitantes, enquanto dois Estados, que deram início ao projeto em 1965, selecionaram unidades de 2.000-2.500. Os dez Estados que começaram as operações em 1966 usaram uma combinação de unidades de 2.000 e de 8.000. Registradores, em expediente parcial, foram utilizados nas unidades menores e outros, em horário integral, foram empregados nas unidades maiores. Em ambos os casos, entretanto, verificou-se que as unidades eram demasiadamente grandes para cobertura adequada. Por isso, foram escolhidas áreas, contendo 750 — 1.000 habitantes, como unidades de amostra para operações em escala completa. Quarteirões de coleta censitária foram geralmente usados como base de amostragem dentro de grandes cidades. Quando não se dispunha destes dados, foram substituídos pelos fundamentos utilizados pelos programas da pesquisa Nacional por Amostragem ou pelos de combate à malária.

Dentro de cada unidade de amostra, a população exposta ao risco são todos os residentes usuais não institucionalizados. Durante o curso da pesquisa e do processo de registro contínuo, as informações são coletadas acerca dos visitantes, bem como dos residentes usuais mas, conforme será comentado mais amplamente em capítulos subsequentes, estão sendo tomados cuidados para distinguir residentes usuais de visitantes temporários e a análise restringe-se aos primeiros.

As pessoas que se mudam para uma unidade de amostra (imigrantes) são consideradas visitantes, até depois da realização do inquérito semi-anual seguinte. Subseqüentemente, serão incluídos na população residente usual. Pessoas que se mudam para fora das unidades de amostra (emigrantes) continuam constando da população-base até que o inquérito seguinte, depois de sua partida, tenha sido efetuado; os eventos desta população são contados até a data de sua migração (Índia, Vital Statistics Division, 1972a).

O plano de amostra do sistema dual de registro indiano prevê a substituição das unidades de amostra (nas áreas urbanas e rurais), quando estas se tornam despovoadas, inacessíveis, ou de outra forma operacionalmente difíceis. Em todos os casos, a seleção de unidades substitutas de amostra foi feita com a colaboração da Repartição do Registro Geral. A experiência mostrou até agora que a substituição de unidades de amostra muitas vezes leva a dados insatisfatórios, pelo menos até que o sistema esteja bem implantado. Assim, por exemplo, 2 unidades, que em 1969 apresentaram coeficientes de natalidade de 33,1 e 32,8, foram substituídas por causa de enchentes; as unidades de substituição apresentaram coeficientes de natalidade de 29,0 e 27,4 para o

primeiro e segundo períodos de 6 meses, respectivamente. Os coeficientes de mortalidade apareceram mais deficientes ainda (Índia, Vital Statistics Division, 1972a).

Desde sua seleção inicial para as bases do censo por amostragem, de 1961, algumas das unidades de amostra sofreram um substancial aumento ou decréscimo no tamanho da população, desordenando assim a classificação original segundo os grupos (size groups). Além disso, algumas das unidades de amostra ultrapassaram o limite dos 2.000 habitantes e tornou-se, portanto, necessária uma subdivisão, ou então uma equipe adicional de registradores. A eficácia das bases do censo de 1971 suscitou as questões:

1) da seleção de novas amostras ou da rotação sistemática das unidades de amostras existentes; ou

2) da conservação da presente amostra, garantindo assim a continuidade da série de coeficientes demográficos. Existem argumentos convincentes para ambas as soluções. Presentemente, sugere-se a manutenção da amostra de 1961 e que a seleção de uma amostra "paralela", comparável em tamanho, seja extraída das bases do censo de 1971. Os resultados desta última podem então ser reciprocamente comparados com a amostra original. A disponibilidade desta amostra paralela pode também aumentar a precisão das estimativas de coeficientes demográficos.

A seleção de nova amostra implicará, entretanto, em maior volume de trabalho operacional e administrativo, incluindo o treinamento de pessoal adicional de campo. Neste sentido, deve-se notar que os atuais planos são de molde a exigir que a amostra paralela seja administrada exclusivamente pelos Escritórios Estaduais do Censo, sem a colaboração do Ministério da Saúde ou o da Economia e Estatística, simplificando assim a estrutura administrativa do sistema.

VII. MAPEAMENTO E OUTROS PROCESSOS INICIAIS

Neste capítulo e nos seguintes, os processos operacionais do sistema dual de registro indiano são descritos detalhadamente. Um organograma destas operações é apresentado na Figura I. A Tabela 2 é um esquema da origem, processamento e destino final de cada um dos formulários usados no sistema. Cópias dos formulários 2, 3, 4, 5 e 6 são reproduzidas nas Figuras 2-6.

Depois da seleção das unidades de amostra, a equipe do escritório central do Estado preparou um mapa apresentando a localização de cada uma delas em relação às divisas distritais, estradas e outros vínculos de comunicação dentro dos limites estaduais. Uma cópia deste mapa foi enviada à Unidade de Registro de Amostra na Repartição Geral do Registro (Índia, Vital Statistics Division, n. d. a.).

O registrador também preparou mapas pormenorizados de cada unidade de amostra. Segundo as instruções, deviam ser preparados esboços da unidade de amostra toda, mostrando, por exemplo, as divisas exatas, todas as rodovias, alamedas, caminhos de carroça, bem como as montanhas e rios e as edificações importantes, tais como: escolas, hospitais e prefeituras. Cada estrutura, inclusive as feitas com material relativamente leve, figurava nestes mapas e era numerada conforme segue.

Nas unidades de amostra, apresentando conjuntos já existentes e relativamente completos de casas numeradas, esse sistema foi mantido e criados números adicionais, quando necessário, através da adição de letras do alfabeto aos números mais próximos em existência (por exemplo: 321, 321A, 322). Onde não havia sistema de numeração utilizável, as casas foram numeradas pelo registrador, adotando qualquer dos planos sistemáticos (números seguindo de nordeste a sudoeste, no sentido dos ponteiros do relógio, ou outro). Cada número de casa, tanto aposto recentemente, como já existente, era marcado no mapa e setas colocadas a intervalos indicavam a ordem ascendente dos números

Os números apostos eram também marcados, a giz ou pixe, em lugar visível, perto da porta de entrada das casas. A fim de facilitar a identificação das unidades residenciais em cada área da amostra, fez-se recentemente um esforço para substituir essas marcas por outras mais duráveis, mas mais dispendiosas: placas de metal.

Quando numeravam as casas, os registradores também faziam entrevistas em cada domicílio visando obter dados para a lista de casas e o levantamento básico. Para o sistema dual de registro, *casa* foi definida mais amplamente como “uma estrutura ou parte de uma estrutura habitada, ou vaga, ou uma moradia, ou uma loja, loja com moradia, ou um recinto comercial, oficina, escola, etc.”. (Índia, Vital Statistics Division, n.d.c., p. 21). A lista de casas continha informações sobre se a casa era usada para fins residenciais ou não-residenciais, se estava vaga ou ocupada, e citava o nome do chefe do domicílio. Quando dois ou mais domicílios ocupavam a mesma casa, cada um era listado separadamente. Para casas com um só domicílio, o número deste coincidia com o da casa; um símbolo alfabético era acrescentado para diferenciar os múltiplos domicílios dentro de uma mesma casa.

O levantamento básico foi efetuado pelo registrador simultaneamente com o processo de numeração e listagem das casas. O programa revisto de domicílios (Formulário 2), apresentado na Figura 2, foi o instrumento principal utilizado no levantamento básico e para o prosseguimento nos inquéritos semi-anuais realizados pelos entrevistadores. Para a série (informação) básica, os registradores preencheram as primeiras sete colunas deste formulário. A parte superior do formulário de levantamento básico contém espaços para anotar o número da casa e do domicílio, o nome do chefe do domicílio, e a religião, bem como os números dos quarteirões e casas, usadas para o censo, sempre

que disponíveis. A parte inferior do formulário destina-se ao nome, grau de parentesco com o chefe do casal, sexo, idade e estado civil de cada membro do domicílio.

Os registradores receberam instruções para alistarem todos os membros de um domicílio, inclusive os temporariamente ausentes, na seguinte ordem: chefe, esposa do chefe, filhas solteiras em ordem cronológica, filhos solteiros em ordem cronológica, filho casado mais velho, sua esposa, seus filhos em ordem de idades, segundo filho casado, sua esposa, seus filhos, e assim por diante. Deu-se ênfase especial à obtenção de informações, referentes a crianças e menores de 1 ano. Outros parentes e empregados foram alistados após os membros da família conjunta. Embora o estudo se referisse à população residente usual, os visitantes temporários também foram alistados no formulário do levantamento básico (ver Figura 2); a coluna 7 foi usada para distinguir *residentes usuais* (inclusive empregados que comem e dormem no emprego, e filhas casadas, cujos maridos estejam no serviço militar ou trabalhando em lugar distante) de *visitantes* (inclusive estudantes que moram longe de suas casas, pessoas que trabalham fora da unidade de amostra, mas regressam nos fins de semana ou feriados, e filhas casadas que voltaram à casa paterna para dar à luz ou para uma breve visita).

A lista de casas, o questionário domiciliar e uma cópia do mapa detalhado permanecem com o registrador na unidade de amostra. São usados pelo supervisor por ocasião das verificações periódicas e são atualizados de 6 em 6 meses pelo entrevistador da pesquisa.

VIII. O PROCESSO DE REGISTRO CONTÍNUO

Quando a operação de mapeamento e levantamento básico estavam completas, os registradores iniciaram sua tarefa principal de observação contínua dos eventos demográficos em cada unidade de amostra. Vários processos são usados para localizar todos os nascimentos e óbitos em cada unidade de amostra. Além disso, as áreas urbanas e rurais requerem tratamentos diferentes: nas unidades rurais, confia-se basicamente num sistema de informantes; nas áreas urbanas, dá-se ênfase especial a freqüentes contatos com todos os domicílios.

Os registradores rurais receberam orientação para estudar os costumes e a estrutura social de suas vilas, a fim de delinear a rede de comunicações de cada uma das localidades ou comunidades dentro da vila. Os registradores rurais deviam também identificar as pessoas que, em virtude de sua posição social, estavam mais aptas a ter conhecimento dos nascimentos e óbitos na população total da unidade de amostra. Inicialmente, foram preparadas listas relativamente longas de contatos em potencial, incluindo pessoas informadas e "tagarelas" em cada localidade ou comunidade, bem como pessoas com probabilidades de terem informações sobre nascimentos e óbitos na vila toda, tais como: parteiras, barbeiros (que muitas vezes desempenham funções es-

pecíficas em conexão com os óbitos), lavadores, e sacerdotes e outros funcionários religiosos. Desta enorme lista de “informantes” iniciais, solicitou-se aos registradores que selecionassem aquelas que fornecessem o máximo de informações sobre ocorrências de nascimentos e óbitos. O objetivo final deste processo foi o de desenvolver uma rede eficiente de informantes e entrar em contato com eles de 15 em 15 dias, ou mais freqüentemente, sempre que possível.

O sistema de informantes usado nos primeiros estágios do plano apresentou certas deficiências na extensão da cobertura dos eventos demográficos. Estas deficiências tornaram-se particularmente evidentes quando ocorria um evento que afetasse um residente usual durante o período de seu afastamento da unidade de amostra, por exemplo, enquanto uma mulher residente estava na casa paterna para dar à luz. Conseqüentemente, pede-se agora aos registradores rurais que visitem trimestralmente cada domicílio em sua unidade de amostra para informar-se acerca de nascimentos e óbitos que, de outra forma, passarão despercebidos.

Nas áreas urbanas, o sistema de informantes não aprovou. Conseqüentemente, os registradores, nas unidades urbanas de amostra, devem agora visitar cada domicílio pelo menos uma vez por mês para saber da ocorrência de nascimentos e óbitos. Para garantir cobertura completa e regular, os registradores urbanos têm que manter um diário mostrando o número de domicílios visitados em cada dia.

Além dos seus contatos com informantes e suas rendas trimestrais ou mensais, tanto os registradores urbanos como os rurais fazem visitas periódicas aos hospitais, maternidades, centros de saúde, cemitérios e crematórios locais. Muitos registradores são professores; seus alunos, por isso, freqüentemente constituem-se em fonte adicional de informações sobre nascimentos e óbitos (Índia, Vital Statistics Division, n.d.c.).

Como outra forma de localizar casos, os registradores preparam listas de mulheres gestantes, baseados em informações colhidas durante suas rondas aos domicílios ou de vizinhos, parteiras, pessoas idosas, centros de assistência à maternidade e à infância. Estas listas foram primeiramente organizadas por ocasião do levantamento básico e são continuamente atualizadas, principalmente pelo registrador (Índia, Vital Statistics Division, n.d.c.).

Ao saber de um nascimento ou óbito, os registradores visitam o domicílio do recém-nascido, ou do falecido, para verificar a ocorrência do evento e obter detalhes. Os instrumentos de coleta de dados, o registro de nascimento e o registro de óbito (Formulários 5 e 6), estão nas Figuras 3 e 4. Estes formulários são preenchidos com todos os eventos demográficos que ocorram com a população residente usual (quer o evento se registre dentro da unidade de amostra ou não) e com todos os visitantes (se o evento se dá dentro da unidade de amostra e chega ao conhecimento do registrador). O ideal é que a informação sobre cada nascimento e óbito seja fornecida pelo chefe do domicílio. Se ele, entretanto, não estiver disponível, outros membros do domicílio podem responder

por ele. Devido à prática do *pardah* (reclusão de mulheres) que predomina em certos segmentos da população indiana, as crianças, às vezes, agem como intermediários, levando aos registradores informações dadas por sua mães.

Na maior parte, os itens nos questionários sobre nascimentos e óbitos são auto-explicativos. Talvez três itens exijam explicação. Os nascimentos e óbitos recebem números seriados seguidos, dentro de cada unidade de amostra, ao serem lançados nos formulários de registro. A coluna 13 nos formulários de nascimento (ver Figura 3) e a coluna 11 no formulário de óbito (ver Figura 4) são usados para registrar se o nascimento ou óbito ocorreu dentro da unidade de amostra:

- 1) a um residente usual presente do domicílio (RUP);
- 2) fora da unidade de amostra a um residente usual do domicílio (RUA, residente usual ausente); ou
- 3) dentro da unidade de amostra a um visitante (V). Eventos nesta última categoria são registrados, mas excluídos para fins de análise. A possibilidade de engano por omissão de um evento com um residente usual é diminuída pela insistência em registrar todos os eventos, quer seja com residentes usuais ou visitantes.

É preciso que os formulários de registro de nascimento e de óbito também indiquem a primeira fonte de informação que levou à identificação do evento. Assim, se o registrador obteve a primeira notícia através de um informante, a ocupação deste é registrada. Anotações adequadas são feitas também nos casos em que as primeiras informações foram: 1) a ronda trimestral dos registradores rurais; 2) a ronda mensal dos registradores urbanos; 3) informações dadas pelo próprio domicílio; ou 4) informações descobertas através de outras fontes. Estes dados constituem a base para avaliação da eficiência dos vários meios utilizados para localizar nascimentos e óbitos, e são usados para aperfeiçoar a eficácia do sistema.

As três colunas-finais (Colunas 19-21) do formulário de nascimento e as Colunas 16 e 17 do formulário de óbito foram introduzidas recentemente para associar os óbitos de menores de 1 ano e de crianças com menos de 3 anos aos registros correspondentes de nascimentos. Quando o nascimento e o óbito ocorrem dentro do mesmo intervalo de 6 meses, o registrador dará a informação correspondente nas Colunas 16 e 17 do formulário de óbitos; nos outros casos, estas colunas são preenchidas no escritório central do Estado depois de associar os dois registros.

No 1.º dia de cada mês, os registradores preparam uma relação mensal de nascimentos e óbitos que é, com efeito, uma cópia de todas as informações lançadas nos registros de nascimento e óbito do mês precedente; estes relatórios são apresentados mesmo nos meses em que não ocorram nascimentos e óbitos. Os registros de nascimento e óbito são

retidos pelo registrador durante o período entre os levantamentos; e são, portanto, registros acumulados abrangendo período de seis meses. São eles submetidos aos escritórios centrais do distrito ou do Estado pouco antes do levantamento semestral, e novos formulários de registro de nascimento e de óbito são então iniciados. Conforme será ventilado no Capítulo XII, os registros de seis meses são usados para confronto com as listas de levantamentos de nascimentos e óbitos.

IX. A ENTREVISTA DO LEVANTAMENTO

Há 3 finalidades básicas para os levantamentos semi-anuais: 1) listar todos os nascimentos e óbitos que ocorram em cada domicílio, bem como os eventos que se verificam com os membros do domicílio ausentes durante um período especificado; 2) atualizar os dados relativos à composição de cada domicílio; e 3) atualizar a lista e o mapa das casas. Para conseguir isso, o entrevistador recebe um mapa, lista de casas, questionários de domicílios, completados durante a *entrevista básica* ou atualizados pelo mais recente levantamento, e a lista de gestantes do registrador e, seguindo a ordem do sistema de numeração de casas, visitará cada casa da unidade de amostra.

Conforme anteriormente mencionado, os registros de nascimento e de óbito, compilados pelos registradores, são recolhidos do campo antes de se iniciar o levantamento semi-anual, de modo que os entrevistadores da pesquisa não têm acesso direto às informações previamente registradas sobre nascimentos e óbitos ocorridos entre os levantamentos.

A lista de casas é atualizada através do acréscimo, no final da lista, de domicílios novos, ou anteriormente omitidos. Cada nova entrada receberá o mesmo número da casa adjacente com a numeração mais baixa. Antepõe-se a letra *N* ou *U* indicando *new* (novo) ou *unnumbered* (não-numerado). Assim, preserva-se a seqüência geográfica do sistema de numeração. Faz-se a marcação adequada no mapa e na lista de casas, entre as duas casas mais próximas, para indicar que foi registrada a casa nova ou recentemente encontrada. Casas demolidas são assim marcadas e os números delas ficarão dentro de um círculo. Anotam-se, também, as alterações no uso residencial e não-residencial de casas anteriormente listadas.

Nos estágios iniciais do projeto, somente alterações nas condições residenciais foram anotadas no questionário de domicílios. Este questionário foi revisto para incluir informações sobre mudanças na faixa etária e no estado conjugal. Os entrevistadores recebem instruções para cuidarem de que as alterações apresentadas sejam compatíveis com as informações coletadas em entrevistas anteriores. Em cada entrevista semi-anual, os entrevistadores introduzem códigos no questionário de domicílios, para diferenciar residentes usuais de visitantes e indicar nascimentos, óbitos, imigrações ou emigrações dentro dos domicílios, e modificações de residência dentro da unidade de amostra. Quando o

entrevistador do levantamento constata que os ocupantes de um domicílio inteiro mudaram de residência, todos os seus membros serão codificados como emigrantes e o questionário do domicílio será levemente cruzado a lápis. Novos questionários serão preenchidos para imigrações de todos os ocupantes de um domicílio.

O questionário de domicílio destina-se a cobrir um período de três anos, inclusive a entrevista básica e seis entrevistas semi-anuais a serem realizadas pelos entrevistadores do levantamento. No fim de três anos os dados são transferidos para novos questionários e o processo continua.

Adicionalmente à coleta de dados referentes às características básicas da população e aos eventos demográficos, os levantamentos periódicos do sistema dual de registro indiano têm sido e continuarão sendo usados para estudos especiais. Por exemplo, depois de um estudo piloto em seis Estados, fez-se uma pesquisa de planejamento familiar numa subamostra abrangendo 10% das unidades totais incluídas no sistema dual de registro. Dentro das unidades designadas da sub-amostra, todos os homens, que tinham mulheres com menos de 45 anos, foram interrogados sobre seus conhecimentos e práticas de métodos de planejamento familiar.

Outra entrevista de levantamento (tendo 1.º de julho de 1972 como data de referência) incluiu, também em base de subamostra, um questionário bastante extenso sobre fecundidade. Continua itens tais como: idade no casamento efetivo, número total de filhos nascidos vivos, e idade da mãe nos nascimentos sucessivos. Questões sobre variáveis econômicas e sociais também foram incluídas, por exemplo, grau de instrução e ocupações de ambos os cônjuges, e despesas domiciliares mensais. Espera-se que os resultados desta pesquisa estejam disponíveis em fins de 1973. Os dados adicionais obtidos assim através destes levantamentos desdobrados ilustram a flexibilidade potencial do sistema dual de registro indiano.

Os eventos demográficos localizados pelos entrevistadores, e as características destes eventos, são lançados em dois formulários adicionais do levantamento, as listas de nascimentos e óbitos (Formulários 3 e 4), mostrados nos Modelos 5 e 6, e cada evento recebe um número seriado consecutivo. Como se pode verificar, as informações sobre nascimentos e óbitos colhidas durante a entrevista do levantamento são menos detalhadas do que as coletadas pelos registradores. Originalmente, a duração do período de referência para a informação retrospectiva de eventos demográficos foi de seis meses. Recentemente foi introduzido o uso experimental de um período de um ano, os entrevistadores do levantamento preenchem listas separadas de levantamento de nascimentos e óbitos para a primeira e segunda metades do período. Este tratamento experimental proporciona oportunidades para o estudo dos efeitos de um período mais longo *para nova visita* quanto a lapsos de memória e fornece também uma base para posteriores confrontos de registros. Por outro lado, o período prolongado para nova visita pode

causar problemas de maior volume de trabalho para os supervisores bem como problemas de comparação e de verificação de campo (Agrawal, 1972 a).

Como outra parte de suas atividades de levantamento, solicitou-se aos entrevistadores que atualizassem a lista de atuais gestantes. Esta lista é depois devolvida ao entrevistador, no fim de cada visita do levantamento, para posterior acompanhamento e atualização através de informações dadas por parteiras ou vizinhos.

Depois de completar as visitas do levantamento, os entrevistadores tabulam os dados relativos à composição de domicílios e preparam um tabela-padrão apresentando a população, segundo o sexo e grupos etários quinquenais (apresentada como Formulário 12 no Modelo 1 e na Tabela 2). Um formulário adicional, chamado cartão da unidade demográfica (apresentando o número de habitantes no levantamento anterior e no atual, bem como o total de nascimentos, óbitos, imigrantes e emigrantes) está sendo atualmente usado em caráter experimental. Este formulário possibilitaria a aplicação da equação comparativa dentro dos limites de cada unidade de amostra e serviria como verificação parcial do controle de qualidade.

Quando o levantamento e a tabulação estiverem completos, o entrevistador devolverá ao registrador o mapa atualizado, a lista de casas, os questionários de domicílios, e a lista de gestantes e enviará a lista dos nascimentos e óbitos coletados durante o levantamento à repartição adequada (sedes estaduais ou distritais) para confronto com os relatórios do registrador. Os formulários que apresentam a distribuição da população, segundo a idade e o sexo, são também mandados às sedes estaduais ou distritais onde serão retidos até que tenham sido completadas as correções, baseadas no resultado da verificação de campo. Uma cópia das tabelas de população será então expedida para as sedes nacionais (repartições centrais).

X. CIRCULAÇÃO DE DOCUMENTOS E CRÍTICA INICIAL

Conforme citado anteriormente, os registradores remetem os relatórios mensais de nascimentos e óbitos às repartições (sedes) estaduais no primeiro dia de cada mês. A data de seu recebimento é registrada num mapa de controle apresentando cada unidade de amostra e cada mês. Lembretes são enviados aos registradores cujos relatórios não foram recebidos até o dia 10 de cada mês; depois de 15 dias é feita a complementação através das repartições (sedes) estaduais ou distritais. Os relatórios mensais são revistos logo que recebidos. Quaisquer discrepâncias percebidas são levadas ao conhecimento dos registradores por correspondência; a repartição estadual também se comunicará diretamente com os registradores com relação aos melhores informantes aos diversos métodos de aperfeiçoamento do trabalho.

Para cada unidade de amostra, a necessidade de mais controles de qualidade, em nível nacional ou estadual, incluem a manutenção, ou de gráficos mostrando as flutuações de mês a mês nos números de nascimentos e óbitos recebidos ou de um formulário de controle comparando eventos demográficos atuais e esperados. Variações anormais nas tendências ou afastamentos acentuados de expectativas são investigados.

O recebimento dos resultados da entrevista do levantamento é também inspecionado através de um gráfico de controle mantido na sede (repartição central) estadual ou distrital. Os resultados são revistos, logo que recebidos, para verificar se estão completos e consistentes. Nos capítulos seguintes, são descritos os processos de correção e a circulação dos documentos durante as fases de comparação e verificação de campo.

XI. COMPARAÇÃO E VARIAÇÃO DE CAMPO

Dependendo da estrutura específica de organização do Estado, a comparação dos relatórios de eventos do registrador e do entrevistador é executada numa de suas sedes (órgãos centrais): estadual, regional, ou distrital. De preferência, esta tarefa caberá a um funcionário, que não seja o entrevistador, nem o registrador, que executaram o trabalho em determinada unidade, e será supervisionada por um funcionário Sênior do Programa. Os registros de nascimentos e óbitos (Formulários 5 e 6) de cada unidade de amostra, completados pelo registrador, serão comparados com as listas de nascimentos e óbitos (Formulários 3 e 4) compiladas pelo entrevistador do levantamento, tão logo estas últimas sejam recebidas. Um evento demográfico é considerado certo se houver correspondência exata entre os dois tipos de registro, com respeito a localização do domicílio (quer seja número domicílio ou nome do seu chefe, ou ambos); nome das mães em casos de nascimentos e óbitos de menores de 1 ano e do morto, em casos de falecimento; situação residencial; sexo; e mês da ocorrência. Quando todos estes 5 itens estão em concordância, assinala-se o número de série que aparece nos registros de nascimento e de óbito; este número de série é então transferido para a Coluna 9 das listas de nascimentos e óbitos, e o código M para "matched" (confrontados) será também lançado na Coluna 9.

Quando todos os eventos plenamente confrontados foram identificados, os eventos remanescentes serão selecionados para confrontos parciais. Para ser considerado um confronto parcial, o número do domicílio, o nome da mãe, ou o do falecido, devem concordar, e, adicionalmente, pelo menos um dos outros três critérios de confronto (sexo, mês da ocorrência, e situação residencial) deve combinar. Quando um confronto parcial é estabelecido, critérios discrepantes recebem um círculo vermelho; o número de série do registro é transferido para a Coluna 9 do formulário de levantamento de nascimentos e óbitos (Formulários 3 e 4) e o código PM (partially matched) (parcialmente con-

frontado) é lançado. Os eventos remanescentes dos registros de nascimentos e de óbitos (que não combinam) são transferidos para o final das listas de levantamento; de nascimentos e óbitos; todos os eventos não confrontados de ambas as fontes são codificados com EX na Coluna 9.

Esses critérios de confronto abrangem virtualmente todos os itens de informação que se podem obter da lista de levantamentos de nascimentos e óbitos (Formulários 3 e 4); sua adequação tem sido testada em parte, através de experimentos feitos no Estado de Gujarat (Mehta, 1970). Adicionalmente, foi efetuado um experimento com 640 nascimentos e 194 óbitos, em Delhi, no qual todos os itens abrangidos pelo registro de nascimentos e óbitos foram incluídos no formulário do levantamento. Quando todos os 17 itens relativos a nascimentos e os 13 referentes a óbitos foram usados no estudo de Delhi, para estabelecer confrontos, somente cerca de um quarto de todos os nascimentos e um quinto de todos os óbitos foram inicialmente classificados como confrontados. Quando os critérios de confronto foram restringidos aos atualmente utilizados em operações de campo, 51 por cento de nascimentos e 28 por cento de óbitos puderam ser completamente confrontados no estágio inicial (Agrawal et al., 1968). A grande proporção de confrontos parciais sobrecarrega em muito o processo de verificação de campo. Portanto, os critérios de confronto em uso corrente parecem ser práticos, embora talvez não sejam ótimos. Planeja-se pesquisa mais detalhada sobre critérios de confronto.

Todos os eventos parcialmente confrontados são devolvidos ao campo para verificação, preferencialmente por uma terceira turma, ou então pelo entrevistador e o registrador, que revisitam os domicílios onde ocorreram os eventos não confrontados ou parcialmente confrontados. O objetivo é determinar se um ou mais relatórios se referem a eventos ocorridos dentro dos limites temporais e geográficos fixados. Os eventos não confrontados são igualmente conferidos para verificar a autenticidade e exatidão dos detalhes; estão sendo envidados esforços no sentido de constatar se alguns dos eventos não-confrontados, realmente constituem eventos.

Ocasionalmente, o processo de verificação de campo revela eventos demográficos adicionais omitidos pelo registrador e pelo entrevistador. Tais eventos são registrados no formulário do registro mensal (que aparece como Formulário 10 no Modelo 1 e Tabela 2), a fonte da primeira informação é listada como verificação de campo. Eventos verificados, só tomados pelos entrevistadores, são também listados num Formulário 10 em separado. Portanto, estes comprovantes, apanhados coletivamente, contém todas as informações disponíveis sobre nascimentos e óbitos.

Depois de concluído o processo de verificação de campo, a lista de levantamento de nascimentos e óbitos, o questionário de domicílios e a distribuição da população segundo a idade e o sexo, são corrigidos à luz dos dados obtidos nessa verificação. Estes formulários são então

submetidos ao centro estadual, onde serão usados para corrigir os relatórios mensais do registrador. Os resultados do confronto e da verificação de campo são compilados num formulário-resumo (Formulário 11 da Tabela 2) apresentando, para cada unidade de amostra, o número de nascimentos e óbitos segundo a fonte, i.é., recolhidos tanto pelo registrador como pelo entrevistador, somente pelo registrador, somente pelo entrevistador, ou de outras fontes.

As taxas de confronto encontradas no sistema dual de registro indiano têm sido, em geral, bastante elevadas; cerca de 80 a 83 por cento dos eventos rurais, e de 70 a 76 por cento dos eventos urbanos identificados foram encontrados em ambos os sistemas (ver Tabela 3). Estas taxas elevadas de confronto podem ter resultado da ênfase colocada na obtenção de informações completas sobre todos os nascimentos e óbitos do sistema de registro e da entrevista do levantamento. A estreita correspondência entre eventos encontrados no levantamento e no processo de registro pode, alternativamente, refletir dependência, inadvertida ou não, entre os dois sistemas.

Segundo indicam os dados na Tabela 3, a atuação dos entrevistadores do levantamento foi, em geral, superior à dos registradores, na identificação dos nascimentos e óbitos. Caso haja convivência entre registradores e entrevistadores, provavelmente tenderá a refletir taxas de confronto mais elevadas para os entrevistadores. A análise dos modelos de taxas de confronto indicou que os registradores estão mais sujeitos a omitir eventos, ocorridos com residentes usuais fora da área da unidade de amostra, do que os entrevistadores (Ramabhadran e Agrawal, 1972). Conforme anteriormente citado, a série trimestral de visitas domiciliares empreendida pelos registradores rurais representa um esforço para superar esse tipo de omissão. Nas unidades de amostras rurais esparsas, os registradores deixaram passar alguns eventos, em consequência de conceitos errôneos sobre os limites exatos da área. Em raros casos, os registradores residem a alguma distância de suas unidades de amostra, e isto também parece contribuir para informações incompletas. A confusão para se determinar o momento exato dos eventos em relação à data de referência também resultou na omissão de alguns eventos vitais.

Até certo ponto, a atuação menos adequada dos registradores, com relação aos entrevistadores, em localizar nascimentos e óbitos, parece ser resultado da incapacidade dos registradores em executar todas as tarefas que lhes foram atribuídas. Assim, por exemplo, verificou-se que alguns registradores rurais confiavam unicamente no sistema de informantes. Em outros casos, os entrevistadores, em vez de entrarem em contato com os informantes, esperam que estes lhes levem as informações. Adicionalmente, parece que alguns registradores de áreas urbanas não conseguiram realizar as visitas mensais de rotina. A supervisão mais intensiva, comentada na Seção IV, deverá contribuir para a solução destas dificuldades.

Os registradores, bem como os entrevistadores do levantamento, têm uma tendência a deixar passar óbitos de infantes e eventos ocorridos em domicílios de um membro, ou de emigrantes, mas, devido ao período de referências mais longo nos levantamentos e à falta de oportunidade em acompanhar as modificações diariamente, estes tipos de omissões parecem ser mais característicos da entrevista do levantamento do que da operação de registro.

Finalmente, um número de fatores associados à falta de evento é inerente às crenças, costumes e atitudes da população inquirida. Já foi mencionado o fato de que, quando os membros do sexo masculino de um domicílio estão ausentes, as crianças, às vezes, atuam como intermediários entre os membros do sexo feminino do domicílio e os registradores e entrevistadores. Adicionalmente, pessoal do serviço de campo comunicou que, em certos domicílios, teme-se a imposição de planejamento familiar, ou a vacinação de infantes, o que provoca a relutância em fornecer informações sobre nascimentos.

Um novo levantamento intensivo, realizado numa amostra aleatória de unidades, no Estado de Kerala, como parte dos processos de verificação de campo, revelou números consideráveis de eventos vitais perfazendo 8 por cento do total de nascimentos e 5 por cento do total de óbitos omitidos por ambos, registradores e entrevistadores. Estes eventos omitidos "observados" são bem mais substanciais do que o número estimado de eventos omitidos, admitindo-se a hipótese de independência pela fórmula Chandra Sekar-Deming (Kurup, 1968).

XII. TABULAÇÃO E ANÁLISE

Assim como o processo de coleta de dados, também a tabulação e análise são efetuados em vários níveis e por várias repartições. A maior parte da tabulação dos dados é feita manualmente na unidade de amostra; os planos distritais e estaduais, e os cálculos, são executados em calculadoras de mesa. Entretanto, a Unidade de Registro de Amostras, na Repartição do Registro Geral, dispõe de um computador IBM 1401.

O processo tem início em nível de amostra, com a tabulação da população, segundo o sexo e a idade (apresentada no Formulário 1: Modelo 1 e Tabela 2), pelo entrevistador do levantamento. Depois das correções, baseadas nos resultados da verificação de campo, os dados para as unidades individuais de amostra são combinados em centros estaduais, para dar a população total, segundo o sexo e a idade, de cada faixa dentro dos limites de cada Estado.

Depois de concluída a verificação de campo, as listas de levantamento de nascimentos e óbitos (Formulários 3 e 4) constituem um contagem não-duplicada completa de todos os eventos encontrados pelos registradores e entrevistadores, como o são os formulários corrigidos dos relatórios mensais (apresentados como Formulário 10 no Modelo 1 e Tabela 2). Espera-se que os Formulários 10 e 12 cheguem a

Escritório do Registro Geral dentro de 4 meses, a partir do início de determinada série de visitas.

Cada Estado efetua uma análise substancial e metodológica. O plano de tabulação mínima para cada Estado é apresentado na Tabela 4. Adicionalmente a estas tabulações, os Estados são incentivados a realizar outras investigações relacionadas com a eficiência das operações e outros estudos analíticos. Planos detalhados para dirigir estes estudos foram encaminhados aos Estados. Cada Estado deverá preparar um relatório anual sobre a operação e os resultados do sistema. As conclusões dos diversos estudos especiais são publicadas no "Sample Registration Bulletin" (Boletim do Registro por Amostra).

Foram feitas análises da mortalidade infantil, das taxas específicas de fecundidade e mortalidade, segundo a idade nos meios rurais, e das flutuações de amostragem (Índia, Vital Statistics Division, 1971a, 1972b), e estão sendo preparados estudos sobre diferenciais de fecundidade segundo aspectos educacionais, religiosos e de parturição, sobre as razões de masculinidade e variações estacionais dos nascimentos e óbitos.

XIII. RELACIONAMENTO DO SISTEMA DUAL DE REGISTRO INDIANO COM OUTRAS ATIVIDADES DE COLETA DE DADOS

Embora o objetivo básico do sistema dual de registro indiano seja o de proporcionar estimativas fidedignas de taxas de natalidade e mortalidade, os dados produzidos pelo sistema apresentam ampla variedade de utilizações potenciais, particularmente na avaliação da fidedignidade dos dados coletados em outros estudos. As investigações conduzidas em colaboração com a Pesquisa Nacional por Amostragem (NSS) e os estudos de verificação de registro realizados com o Censo de 1971 proporcionam exemplos da utilização potencial do sistema neste sentido.

Um dos estudos comparativos da Pesquisa Nacional por Amostragem (NSS) e do sistema dual de registro foi realizado na Bengala Ocidental, durante o período 1967-68. A pesquisa (NSS) tomou 20 domicílios de cada uma das 92 unidades de amostra abrangidas pelo sistema dual de registro. A entrevista da Pesquisa Nacional por amostragem solicitou informações retrospectivas sobre nascimentos e óbitos ocorridos no período anterior de 2 anos; os resultados foram tabulados para mostrar os nascimentos e óbitos ocorridos no período anterior à pesquisa. A comparação dos eventos vitais encontrados pela NSS com os resultados do sistema dual de registro provou que neste último, os eventos eram relatados mais completamente e as datas eram mais precisas. Um segundo estudo comparativo foi efetuado em 57 unidades de amostra, em Tamil-Nadu, em 1968-69; neste estudo, unidades de amos-

tra inteiras, em vez de subamostras, foram cobertas pela NSS. Os resultados deste estudo estão sendo analisados atualmente.

Duas verificações cruzadas com os resultados do Censo de 1971 estão atualmente em andamento. Numa delas, dados do sistema dual de registro indiano estão sendo usados para estimar a integridade da resposta a uma pergunta retrospectiva do censo sobre nascimentos vindos de mulheres atualmente casadas durante o ano anterior ao censo. O segundo estudo refere-se à mensuração da amplitude da subcoleta censitária de menores na faixa etária de 0 a 4 anos. Esta investigação pesquisa os sobreviventes de nascimentos registrados no sistema dual de registro e compara-os com a população de 0, 1, 2, etc. anos de idade registrados pelo censo.

Todos estes estudos deverão levar a aperfeiçoamentos, não somente do sistema dual de registro, mas também de outros sistemas estatísticos dentro da Índia.

XIV. ESTUDOS ANALÍTICOS E DE AVALIAÇÃO

Embora tenham sido feitas referências aos resultados de diversos estudos analíticos e de avaliação, em seções anteriores, o esquema sistemático dentro do qual estas investigações foram executadas não foi convenientemente discutido. Em vista do grande volume de dados produzidos pelo sistema, e da ênfase dada à manutenção de elevados padrões de qualidade, o planejamento para ambas, a pesquisa essencial e a avaliativa, é difícil. Uma das principais forças do sistema dual de registro indiano é a amplitude e profundidade de seu planejamento de pesquisa; por isso, parece oportuno que neste capítulo se apresente uma visão mais integrada do programa analítico e de avaliação do sistema dual de registro.

As principais áreas de pesquisa foram classificadas sob os seguintes cabeçalhos:

1. Eficiência do plano de amostragem
2. Desempenho dos registradores e entrevistadores de pesquisa
3. Confronto (coincidências)
4. Coeficientes vitais
5. Mortalidade infantil e coeficientes afins
6. População
7. Migração
8. Diversos.

Conforme anteriormente mencionado, a pesquisa da adequação do plano existente de amostragem tornou-se imediata em consequência da disponibilidade dos fundamentos da amostragem do Censo de 1971. Assuntos correlatos referem-se ao nível de precisão das estimativas, ao tamanho ótimo das unidades de amostra, aos métodos de segmentação de povoados maiores, aos processos de estratificação, à alternância e à reposição de unidades de amostra.

A eficiência das operações de campo está sendo examinada por uma variedade de processos, alguns dos quais foram sintetizados em capítulos anteriores. O desempenho relativo de registradores e agentes censitários é investigado através de comparações de coeficientes de compreensibilidade de ambos os componentes do sistema. Pesquisa deste tipo pode ser ampliada para incluir o estudo dos efeitos das características do registrador e do entrevistador (idade, ocupação, etc.) sobre os coeficientes de confronto. Esta forma de estudo foi suplementada pela aplicação de entrevistas intensivas por equipe mais qualificada, num esforço de localizar eventos omitidos pelos registradores e pelos entrevistadores da pesquisa (O estudo feito em Kerala, anteriormente citado, é um exemplo deste tipo de investigação). O estudo das características de eventos omitidos pelos registradores, pelos entrevistadores, ou — quando os dados são obtidos de entrevistas intensivas — por ambos os subsistemas, abrange outra área de pesquisa; os resultados destes estudos são de valor evidente na identificação das dificuldades específicas de ambos os componentes dos sistemas dual de registro. A relativa eficiência das diversas fontes de informação (informantes, rondas de domicílio a domicílio, etc.) também está sendo estudada.

O desenvolvimento e a avaliação preliminar dos critérios de confronto também foram descritos em capítulos anteriores. Tanto as investigações sintéticas, como as empíricas, do efeito das estratégias de confronto alternado, demonstraram que o rigor dos critérios de confronto pode ter um efeito substancial sobre a estimativa de coeficientes vitais. Daí a ênfase para uma pesquisa adicional sobre os processos de confronto e verificação e campo.

Dados substanciais e estudos analíticos concernentes às categorias 4-8, abrangendo resultados básicos da nação como um todo, de cada Estado e de cada estrato, foram publicados numa série de relatórios gerais e também no *Sample Registration Bulletin*. Tendo em vista o objetivo primordial de medir os coeficientes vitais, a maioria dos relatórios substanciais de dados dedicou-se às tendências e diferenciais na fecundidade e na mortalidade. Como se reconhece gradativamente, o sistema dual de registro indiano produz dados em ampla variedade de características sociais e demográficas, alargando-se conseqüentemente o conteúdo dos relatórios. A análise projetada tratará de tópicos, tais como: a idade no casamento efetivo, composição do domicílio e migração.

XV. FONTES DE REFERÊNCIAS

TABELA 4

PLANO DE TABULAÇÃO MÍNIMA PARA OS ESTADOS: SISTEMA DUAL DE REGISTRO INDIANO LISTA DE TABELAS

A. População

1. Número de unidades, população de 1961 no universo e na amostra por estrato.
2. População estimada por estrato, em meados do ano.
3. População, segundo o sexo e grupos de idade (total das unidades de amostra), em meados do ano, pela divisão natural, classes de cidades de acordo com o número de habitantes.
4. População, em meados do ano, e população de 1961, segundo o sexo e por unidade de amostra.

B. Natalidade

1. Nascidos vivos, segundo o sexo, em unidades de amostra, total estimado de nascidos vivos e coeficiente estimado de nascimentos por estrato.
2. Nascidos vivos, segundo o sexo e mês da ocorrência, em cada unidade de amostra.
3. Taxa de fecundidade segundo a idade.
4. Nascidos vivos, segundo o tipo de atendimento no parto (total das unidades de amostra) em cada estrato.

C. Mortalidade

1. Óbitos, segundo o sexo, nas unidades de amostra e estimativas de óbitos e de taxa de mortalidade por estrato.
2. Óbitos, segundo os grupos de idade e o sexo (total das unidades de amostra), por divisão natural/cidades segundo o tamanho.
3. Taxa de mortalidade segundo a idade.
4. Mortalidade, segundo o sexo e o mês da ocorrência, por unidade de amostra.
5. Mortalidade, segundo o tipo de atendimento.

D. Mortalidade Infantil

1. Mortalidade infantil, segundo o sexo, em unidades de amostra, estimativas de mortalidade infantil e de taxas de mortalidade infantil por estrato.

E. Geral

1. Estimativas de População, de Nascidos Vivos, de Coeficientes de Nascidos Vivos, de Mortalidade Infantil, de Óbitos, de Taxas de Mortalidade, e de Taxas de Mortalidade Infantil, para cada Levantamento Semi-anual.
2. Comparação de taxas com outras fontes (Pró-forma, segundo as informações disponíveis).
3. Nascidos vivos/óbitos registrados pelo registrador e entrevistador da pesquisa, por estrato.
4. Eventos extras verificados na visita trimestral e período extenso de referências, de 1 ano, por unidade. (Pró-forma para visitas trimestrais; informações podem ser projetadas por Estado).

FONTE: Vital Statistics Division 1972 Sample Registration of Births and Deaths in India, 1969-70 New Delhi: Office of the Registrar General, p. 137

TABELA 1

ÓRGÃO EXECUTANTE: TAMANHO DA AMOSTRA E DATA DO INÍCIO
DAS OPERAÇÕES COLETADAS: SISTEMA DUAL
DE REGISTRO INDIANO

ESTADO/TERRITÓRIO FEDERAL 1	ÓRGÃO EXECUTANTE		Nº DE UNIDADE DE AMOSTRA		DATA DO INÍCIO	
	Rural 2	Urbano 3	Rural 4	Urbano 5	Rural 5	Urbano 7
A ESTADOS						
1 — Andhra Pradesh	Saúde	DCO	150	60	1- 1-68	1-7-69
2 — Assam	DCO	DCO	100	60	1- 7-67	1-7-68
3 — Bihar	E & S	E & S	150	60	1- 7-66	1-1-69
4 — (a) Gujarat	Saúde	DCO	136	60	1-10-65	1-7-69
(b) Maira	Saúde	Saúde	30	16	1- 9-65	1-8-65
5 — Haryana	Saúde	Saúde	60	60	1- 7-68	1-7-69
6 — Himachal Pradesh	DCO	DCO	40	20+20=40 (1-1-70)	1- 1-70	1-7-69
7 — Jammu & Kashmir	DCO	DCO	50+30=80 (1-1-70)	60	1- 7-67	1-7-68
8 — Kerala	E & S	DCO	150	60	1- 7-65	1-1-70
9 — Madhya Pradesh	E & S	DCO	150	60	1- 1-70	1-7-69
10 — Maharashtra	E & S	DCO	150	100	1- 7-65	1-7-69
11 — Mysore	E & S	E & S	150	60	1- 6-66	1-7-68
12 — Nagaland	DCO	—	25	—	—	—
13 — Orissa	Saúde	Saúde	150	60	1- 1-68	1-1-69
14 — Punjab	Saúde	Saúde	80	60	1- 1-68	1-1-69
15 — Rajasthan	E & S	E & S	150	60	1- 7-67	1-1-69
16 — Tamil Nadu	Saúde	DCO	150	100	1- 7-68	1-7-69
17 — Uttar Pradesh	DCO	DCO	100+50=150 (1-1-70)	100	1- 7-67	1-7-68
18 — West Bengal	Saúde	DCO	150	100	1- 1-68	1-7-69
			2 201	1 176		
B TERRITÓRIOS FEDERAIS						
1 — Andaman & Nicobar Islands	DCO	DCO	20	4	1- 1-71	1-1-71
2 — Chandigarh	DCO	DCO	5	10	1- 7-71	1-7-71
3 — Dadra & Nagar Haveli	DCO (Gujarat)	—	10	—	1- 1-71	—
4 — Delhi	DCO	DCO	20+20=40 (1-1-70)	40	1- 8-67	1-1-68
5 — (a) Goa	E & S	E & S	40	20	1- 1-70	1-1-70
(b) Daman & Diu	DCO (Gujarat)	DCO (Gujarat)	5	4	1- 1-71	1-1-71
6 — L M & A Islands	DCO (LMA)	—	5	—	1- 1-71	—
7 — Manipur	DCO	DCO	26	10	1- 7-68	1-7-68
8 — N E F A	DCO	—	26	—	1- 7-70	—
9 — Pondicherry	DCO (Tamil Nadu)	DCO	25	10	1- 1-71	1-1-71
10 — Tripura	DCO	DCO	10	10	1- 1-70	1-1-70
		SOMA:	211	108		
		TOTAL GERAL:	2 412	1 284		

DCO: — Director of Census Operations

E & S: — Directorate/Bureau of Economics and Statistics

* Ainda não iniciado

— Nenhuma área urbana

FONTE: India Vital Statistics Division 1972 Sample Registration of Births and Deaths in India, 1969-70 New Delhi: Office of the Registrar General, p 33

TABELA 2

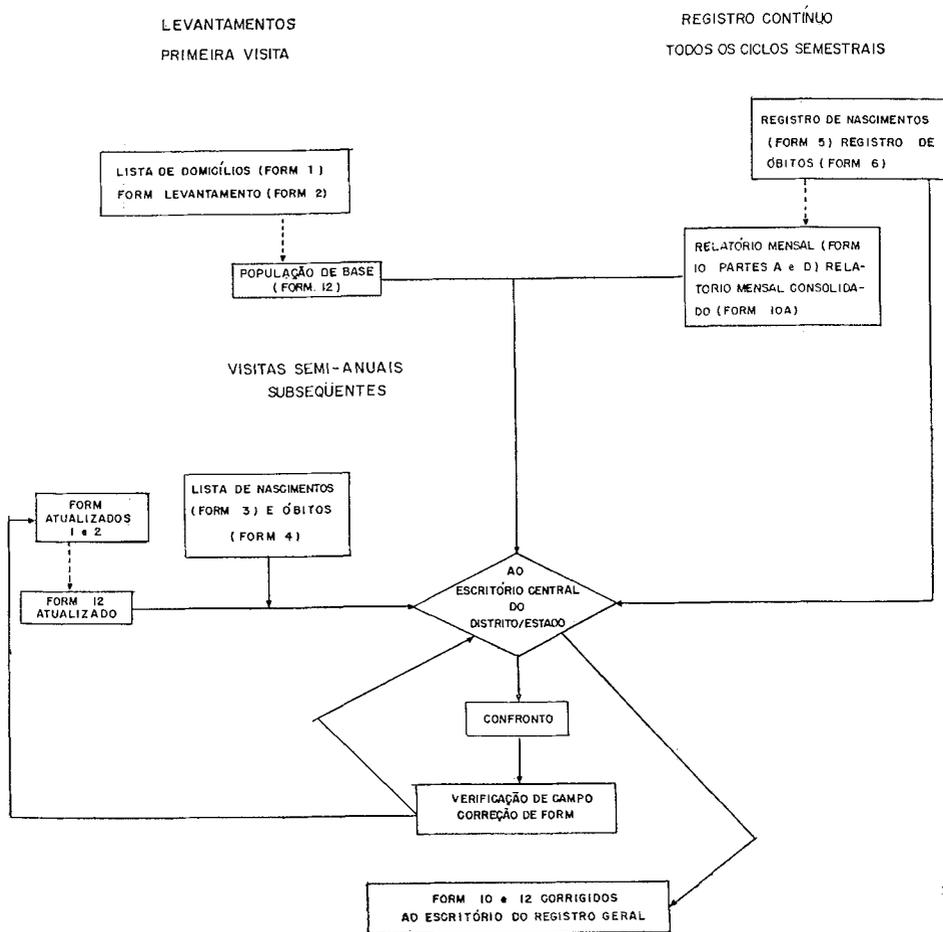
**PREPARAÇÃO E GRÁFICO DE RELATÓRIOS: SISTEMA DUAL
DE REGISTRO INDIANO**

FORMULÁRIO N.º	TÍTULO	PREPARAÇÃO		AÇÃO PÓS PREPARATÓRIA	REMESSA		NOTAS
		Por quem	Quando		Quando	A quem	
1	Lista de casas	Registrador	Pouco depois de regressar do treinamento	Guardado na unidade de amostra para referência e para ser atualizado pelo registrador em cada levantamento sucessivo	—	—	—
2	Questionário de domicílio	Registrador	Pouco depois de regressar do treinamento	Usado pelos entrevistadores para censo de população e para ser atualizado pelo registrador em cada levantamento sucessivo	—	—	—
3	Lista de nascidos vivos	Entrevistador	Durante a pesquisa semi-anual	Confronto do registro de nascimento (Form. 5) no escritório estadual central	Um dia depois de concluído	Estado/escritório/distribuidor	Ao Órgão de Registro Geral quando do solicitado
4	Lista de óbitos	Entrevistador	Durante a pesquisa semi-anual	Confronto com registro de óbitos (Form. 6) no escritório estadual central	Um dia depois de concluído	Estado/escritório/distribuidor	Ao Órgão de Registro Geral quando do solicitado
5	Registro de nascimento	Registrador	Data do registro inicial	Usado para: (a) preparar Relatório Mensal (Form 10); (b) confronto com Form 3 no escritório estadual central	Ao fim de 6 meses e antes de realizar o levantamento semi-anual seguinte	Estado/escritório/distribuidor	—
6	Registro de óbitos	Registrador	Data do registro inicial	Usado para: (a) preparar Relatório Mensal (Form 10); (b) confronto com Form 4 no escritório estadual central	Ao fim de 6 meses e antes de realizar o levantamento semi-anual seguinte	Estado/escritório/distribuidor	—
7	Lista de mulheres gestantes	Registrador	Na ocasião do preenchimento dos questionários domicílio	Continuamente atualizado	—	—	—
8	Relatório de inspeção	(a) Entrevistador; (b) outros funcionários de inspeção	Para a inspeção inicial e outras	Instruções ao registrador para remover falhas	Um dia depois da inspeção	Escritório estadual central	Cópia do Relatório de Inspeção ao Órgão de Registro Geral
10	Relatório Mensal	Registrador	A medida que os eventos são registrados ou no último dia do mês	Reverificação com registros de nascimento e óbito	1º dia do mês	Escritório estadual central	A Repartição do Registro Geral ao término do levantamento semi-anual
10 A	Relatório Mensal consolidado	Escritório estadual central	Dentro de 30 dias	Para garantir que todos os relatórios recebidos pelas unidades foram justificadas	No dia 30 de cada mês	Repartição do Registro Geral	—
11	Resultados de levantamentos semi-anuais	Escritório estadual central	Dentro dos últimos 5 dias ao final do período de 6 meses	Controle completo com Formulários definitivos 10, 3 e 4	Dentro de 4 meses	Repartição do Registro Geral	—
12	População segundo a idade e o sexo	Entrevistador	Durante o levantamento semi-anual	Para controlar totais cruzados e dados do levantamento semi-anual anterior e o n.º de nascimentos, óbitos e migração líquida durante o semestre	No dia da conclusão no campo	Escritório estadual central	Cópia a Repartição do Registro Geral com resultado do levantamento semi-anual
13	Fecundidade	Entrevistador	Durante o levantamento semi-anual	De 3 em 3 anos	No dia seguinte a conclusão	Escritório estadual central	A ser enviado a Repartição do Registro Geral

FONTE: India Vital Statistics Division and Sample Registration of Births and Deaths in India, Neaquarter' Manual. New Delhi: Office of the Registrar General, pp. 34-35

MODELO 1

GRÁFICO DE ROTINA DE OPERAÇÕES: SISTEMA DUAL DE REGISTRO INDIANO



MODELO 2 (CONTINUAÇÃO)

HYS 1- -19			HYS 1- -19			HYS 1- -19			CÓDIGOS
Idade	Estado		Idade	Estado		Idade	Estado		
	Civil	Residencial		Civil	Residencial		Civil	Residencial	
17	18	19	20	21	22	23	24	25	
									U R Residente usual do domicílio, tanto presente como temporariame
									I Imigrado
									O Emigrado
									D Falecido
									N B Recém nascido
									S I Mudado para dentro do domicílio (dentro da Unidade de Amostr
									S O Mudado para fora do domicílio (dentro da Unidade de Amostr)
									V Visitante
									M Casado
									U M Solteiro
									W Viúvo
									D Divorciado
									S Separado
									N B D Recém nascido morto durante o semestre

POPULAÇÃO
RESIDENTE
USUAL

MASCULINO _____
FEMININO _____
TOTAL _____

FONTE: India. Vital Statistics Division 1972. Sample Registration of Births and Deaths in India, 1969-70 New Delhi: Office of the Registrar General, pp 82
*HYS = levantamento semestral (semi anual)

MODELO 3

FORMULÁRIO 5 — REGISTRO DE NASCIMENTOS

POVOADO/CIDADE

BAIRRO

TEHSIL/TALUK

DISTRITO

CÓDIGO DA AMOSTRA

N.º SI.	INFORMAÇÕES SOBRE RECÉM NASCIDOS							CASA N.º E DOMICÍLIO N.º		NOME DO CHEFE (CABECA)	INFORMAÇÕES SOBRE A MÃE QUE DEU A LUZ					TIPO DE ATENDIMENTO NO PARTO	PRIMEIRA FONTE DE INFORMAÇÃO	ASSINATURA DO RECALESEADOR	DATA DO ÓBITO	IDADE DO FALECIDO	NOTAS
	LOCAL DO NASCIMENTO (DENTRO/FORA DA UNIDADE)	DATA DO NASCIMENTO	DATA DA COLETA	NASCIDO VIVOS (NV/NM)	TIPO DE NASCIMENTO SIMPLES/MÚLTIPLO (S/M)	SEXO M/F	EXISTENTES	CENSO DE 1971	NOME DA MÃE		PARENTESCO DA MÃE COM O CHEFE	PADRÃO RESIDENCIAL URP/URA/V	IDADE ATUAL EM ANOS COMPLETOS	RELIGIÃO							
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	

MONTE: India Vital Statistics Division 1972 Sample Registration of Births and Deaths in India, 1969-70 New Delhi: Office of the Registrar General, p 85

MODELO 4

FORMULÁRIO 6 — REGISTRO DE ÓBITOS

POVOADO/CIDADE BAIRRO TEHSIL/TALUK DISTRITO CÓDIGO DA AMOSTRA

N.º SI.	LOCAL DO NASCIMENTO (DENTRO/FORA DA UNIDADE)	DATA DO ÓBITO	DATA DA COLETA	NOME DO FALECIDO	NOME DO PAI/MÃE/ESPOSO DO MORTO	CASA N.º E DOMICÍLIO N.º		NOME DO CHEFE	PARENTESCO DO MORTO COM O CHEFE (COL. 9)	PADRÃO DO MORTO URP/URA/V	SEXO	IDADE AO MORRER			DATA DO NASCIMENTO/MES/ANO QUANDO IDADE AO MORRER E INFERIOR A 3 ANOS	N.º DE SERIE DO NASCIMENTO NO FORMULÁRIO 5/FORMULÁRIO 2	ESTADO CIVIL	RELIGIÃO	ASSISTENCIA MEDICA ANTES DO ÓBITO	PRIMEIRA FONTE DE INFORMACÃO	ASSINATURA DO RELENSEADOR
						EXISTENTES	CENSO DE 1971					MENOS DE 1 MES (EM DIAS)	MENOS DE 1 ANO (EM MESES)	UM ANO E MAIS (EM ANOS COMPLETOS)							
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22

FONTE: Índia Vital Statistics Division 1972 Sample Registration of Births and Deaths in India, 1969-70 New Delhi: Office of the Registrar General p 86

MODELO 5
 FORMULÁRIO 3 — LISTA DE NASCIMENTOS

POVOADO/CIDADE
 TEHSIL/TALUK
 DISTRITO
 CÓDIGO DA AMOSTRA

PERÍODO DE REFERÊNCIA
 (MESES ABRANGIDOS)
 DE A
 PERÍODO DE PESQUISA
 (DATAS DAS PESQUISAS DE CAMPO)
 DE A

N.º SI	CASA N.º E DOMICÍLIO N.º		NOME DO CHEFE DO DOMICÍLIO	INFORMAÇÕES SOBRE A MÃE		INFORMAÇÕES SOBRE FILHOS NASCIDOS			NOTAS DE CONFRONTO	REVERIFICAÇÃO	NOTAS
	Existente	Censo de 1971		Nome	Padrão (URP/URA/V)	Sexo M/F	Data de Nascimento	Nascidos Vivos/ Natimortos (NV/NM)			
1	2	2A	3	4	5	6	7	8	9	10	11

ASSINATURA E DESIGNAÇÃO

FONTE: Índia Vital Statistics Division 1972 Sample Registration of Births and Deaths in India, 1969-70 New Delhi: Office of the Registrar General p 84

MODELO 6

FORMULÁRIO 4 — LISTA DE ÓBITOS

POVOADO/CIDADE
TEHSIL/TALUK
DISTRITO
CÓDIGO DA AMOSTRA

PERÍODO DE REFERÊNCIA
(MESES ABRANGIDOS)
DE A
PERÍODO DE PESQUISA
(DATAS DAS PESQUISAS DE CAMPO)
DE A

N.º SI	CASA N.º E DOMICÍLIO N.º		NOME DO CHEFE DO DOMICÍLIO	INFORMAÇÕES SOBRE O MORTO					NOTAS DE CONFRONTO	REVERIFI-CAÇÃO	NOTAS
	Existente	Censo de 1971		Nome	Padrão (URP/URA/V)	Sexo M/F	Data do Óbito	Idade do Falecido			
1	2	2A	3	4	5	6	7	8	9	10	11

TABELA 3

**OPERAÇÃO COMPARATIVA DE REGISTRADORES E
ENTREVISTADORES DE PESQUISA NO SISTEMA
DUAL DE REGISTRO NA ÍNDIA**

PERÍODO	TOTAL DE EVENTOS COM RESIDENTES USUAIS	DISTRIBUIÇÃO PERCENTUAL				TOTAL % OMITIDO	
		Confrontado	Registro soimente	Entrevista soimente	Omitidos por ambos, mas obser- vados depois	Registrador (cols 5&6)	Entrevistador de Pesquisa (cols 4,6,7)
1	2	3	4	5	6	7	8
URBANO		NASCIDOS VIVOS					
Janeiro ..	13 781	75,8	5,6	16,7	1,9	18,6	7,5
Junho, 70	16 978	76,3	6,1	15,9	1,7	17,6	7,8
Julho ..							
Dezembro, 70	12 993	74,2	6,8	18,6	0,4	19,0	7,2
Janeiro ..							
Junho, 71*							
		ÓBITOS					
Janeiro ..	4 764	72,4	5,9	19,7	2,0	21,7	7,9
Junho, 70	5 354	71,7	6,5	20,1	1,7	21,8	8,2
Julho ..							
Dezembro, 70	4 500	68,9	7,5	23,3	0,3	23,6	7,3
Janeiro ..							
Junho, 71*							
RURAL		NASCIDOS VIVOS					
Janeiro ..	36 494	83,1	5,9	9,8	1,2	11,0	7,1
Junho, 70	43 483	82,6	5,5	10,6	0,9	11,7	8,4
Julho ..							
Dezembro, 70	38 936	81,3	6,5	12,0	0,2	12,2	6,7
Janeiro ..							
Junho, 71*							
		ÓBITOS					
Janeiro ..	15 671	82,8	6,9	8,7	1,6	10,3	8,5
Junho, 70	17 720	81,3	6,4	11,1	1,2	12,3	7,6
Julho ..							
Dezembro, 70	16 560	80,9	7,5	11,4	0,2	11,6	7,7
Janeiro ..							
Junho, 71*							

* PROVISÓRIO

FONTE: Ramabhadran, V.N and Agrawal, R.D 1972 A Preliminary Note on Missing of Vital Events In Indian Sample System. Sample registration Bulletin 6, n° 1: 25-26

37.º ANIVERSÁRIO DO IBGE

O transcurso do “DIA DO ESTATÍSTICO E DO GEÓGRAFO”, festejado no dia 28 de maio último, quando se comemorou o 37.º ANIVERSÁRIO DO IBGE, ensejou diversas solenidades em todo o País.

No Rio de Janeiro, realizou-se o já tradicional “Encontro Ecumênico”, no Clube de Aeronáutica, ocasião em que o Professor Isaac Kerstenetzky, Presidente do IBGE, fez o pronunciamento que a seguir transcrevemos:

“Na esfera da Estatística, as atividades oficiais, no Brasil, remontam a época anterior à realização do I Recenseamento Geral do País, em 1872. Limitavam-se, porém, a inquéritos locais, elementares — mera contagem de habitações e pessoas.

A partir desse levantamento de âmbito nacional, mas adstrito ao aspecto demográfico — como os de 1890, 1900 e 1920, este último incluindo alguns setores econômicos —, os serviços estatísticos governamentais entraram a ter funcionamento regular, supervisionados pela antiga Diretoria-Geral de Es-

tatística, que veio do Império e existiu até pouco depois de 1930.

Desde então, tomaram corpo estudos dos quais resultou a criação, em 1934, do *Instituto Nacional de Estatística*, instalado em 1936, com a finalidade de implantar no País a coordenação dos serviços de estatística nas órbitas federal, estadual e municipal, tendo como órgão deliberativo o *Conselho Nacional de Estatística*.

Foi, assim, da criação do *Instituto Nacional de Estatística* que, em 1938, surgiu o INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA — IBGE, com a reunião do *Conselho Nacional de Geografia* ao *Conselho Nacional de Estatística*.

Três áreas de informações e estudos de grande importância para o desenvolvimento do País estavam abrigadas numa única Instituição — a Cartografia, a Geografia e a Estatística.

Longa foi a evolução e expressivo foi o progresso alcançado em todas as áreas do IBGE durante os 37 anos de sua existência.

Esse progresso tem sido menos o fruto da especialização estanque do que da integração.

É fácil compreender a razão disso quando se percebe que o IBGE constitui um sistema sinérgico. Pois, sem captar-se a interdependência entre as suas partes componentes, o todo não pode ter desenvolvimento adequado.

O Governo teve essa compreensão ao proporcionar à Instituição, nos últimos anos, novo conjunto de instrumentos legais e recursos adequados para sua expansão.

Os grandes projetos e tarefas em que se acha empenhado o IBGE dependem cada vez mais da dedicação plena e entusiástica de seus servidores, bem como da progressiva assimilação de novas gerações. Chego mesmo a divisar, num futuro não muito longínquo, o advento na entidade de uma espécie de "homo ibgeensis" especialista sólido e de visão geral segura a um só tempo.

Passado é o tempo em que o IBGE era mero produtor de dados para uso e elaboração de terceiros. Não que essa função do IBGE não seja de importância fundamental. O planejamento econômico e social, porém, passou a exigir a abertura de novas frentes, e novos desafios surgiram — estatísticas derivadas, abrangendo tanto sistemas de referência complexos como indicadores, pesquisas e análises econômicas e sociais (como o aprofundamento dos estudos dos processos de urbanização e de migração), e estudos vários, inclusive os relativos ao meio ambiente e recursos naturais.

Além disso, passou a ser vital para o IBGE a capacidade de se antecipar a novas solicitações de usuários governamentais, principalmente em razão do longo período de maturação que caracteriza seus projetos mais complexos. O Estudo Nacional da Despesa Familiar, como investigação de múltiplos objetivos é, nesse particular, uma ilustração recente bastante expressiva e talvez a pesquisa social mais significativa já realizada no País.

Como a realidade social é um todo que fracionamos pagando frequentemente preço elevado em termos de relevância, é preciso que, de modo contínuo, tenhamos presente a visão conjunta subjacente das diferentes áreas de competência do IBGE.

Ao comemorarmos, nesta solenidade, o Dia do Estatístico e do Geógrafo, tenho em mente as pessoas assim tituladas menos como membros de corporações medievais e mais como uma legião de dedicados profissionais, com um respeitável acervo de contribuições para o melhor conhecimento da realidade humana e física do nosso País e orientação para seu crescimento e progresso contínuos".

Igualmente destacável foi a cerimônia realizada no Auditório da Escola Nacional de Ciências Estatísticas, promovida pelo Conselho Regional de Estatística (CONRE — 2.^a Região) e a Associação Profissional dos Estatísticos do Brasil (APEB), onde foram debatidos importantes problemas profissionais da classe, culminando com a palestra realizada pelo Professor João Lyra Madeira.

INSTITUTO INTERAMERICANO DE ESTATÍSTICA

O IASI, que é o organismo regional de maior expressão no campo estatístico, cuja fundação se verificou em 12 de maio de 1940, realizou no Panamá a XXXVII Sessão do seu Comitê Executivo, no período de 18 a 24 de fevereiro deste ano.

Na ocasião foram tratados assuntos pertinentes às atividades do Instituto e do Departamento de Estatística da Organização dos Estados Americanos (OEA), salientando-se o exame do relatório anual do Secretário-Geral do Instituto, a situação de coordenação da estatística internacional, as conferências interamericanas de estatística, as assembleias-gerais dos membros do Instituto, a Comissão de Melhoria das Estatísticas Nacionais (COINS), as subcomissões e a Junta Coordenadora, estatuto, regulamento, membros, estado financeiro, etc.

Compõem o Comitê Executivo: Ruben Gleason Galicia (México), Presidente; Raul Sierra Franco (Guatemala), Walter E. Duffet (Canadá), Charles G. Alleyne (Barbados) e Rudolf W. F. Wuensche (Brasil), Vice-Presidentes; e Tulo Hostílio Montenegro (Brasil), Secretário-Geral.

O Instituto Interamericano de Estatística — IASI, tem como finalidade:

a — Estimular o aperfeiçoamento e metodologia do levantamento de dados, preparo de tabulação, análise e publicações de estatísticas, quer oficiais ou extra-oficiais;

b — Promover o emprego de medidas visando melhorar a comparabilidade e o aproveitamento das estatísticas econômicas e sociais entre os países do continente americano;

c — Proporcionar meios para colaboração técnica entre os estatísticos do hemisfério americano e,

d — Cooperar com organismos nacionais e internacionais no progresso da ciência e administração da estatística.

O Conselho da Organização dos Estados Americanos (OEA) e o Instituto Interamericano de Estatística firmaram, em 25 de outubro de 1955, um convênio, mediante o qual o IASI se incorporou aos programas da OEA.

Em consequência, o Departamento de Estatística da Secretaria-Geral da OEA assumiu as funções da Secretaria-Geral do Instituto.

O Convênio atribuiu às Conferências Interamericanas de Estatística o caráter de Conferências Especializadas Interamericanas, de acordo com a Carta da Organização, nas quais se debatem, em alto nível, os problemas estatísticos do Continente e se adotam medidas comuns a todos os países membros. Estas conferências são convocadas pelo Conselho da OEA, tendo a última se realizado em Santiago do Chile, em 1972.

A Comissão de Melhoria das Estatísticas Nacionais (COINS) é um órgão permanente

de consulta do IASI, e é constituída pelos diretores gerais de estatística dos países membros da OEA. Sua convocação se dá pelo Comitê Executivo do IASI, com a finalidade de tratar de aspectos relacionados com os sistemas nacionais de estatística, em busca de soluções adequadas e o estabelecimento de normas interamericanas. A última sessão da COINS foi realizada em Ottawa (Canadá) em 1973, e a próxima terá como sede Lima (Peru), em 1975. Esta Comissão designa, por sua vez, subcomissões e grupos técnicos de trabalho. Uma Junta Coordenadora composta de cinco membros representa a Comissão entre uma e outra sessão e serve, ainda, como órgão consultivo, ao Comitê Executivo e à Secretaria-Geral do IASI.

Presentemente, o IASI conta com 518 associados que, por disposição estatutária, são cidadãos

residentes em países americanos, profissionais de estatística ou vinculados profissionalmente a este campo.

O Instituto, vem trabalhando, nos últimos anos, ativamente no melhoramento das estatísticas nacionais, no desenvolvimento e promoção de estatísticas contínuas (Programa Interamericano de Estatísticas Básicas — PIEB), na realização de censos nacionais como parte do Programa do Censo da América, na assessoria técnica direta, no treinamento em estatísticas básicas e especializadas em nível nacional e internacional, e na reunião e divulgação de material estatístico dos países, através de seu ponto focal de informação. As atividades do Instituto são divulgadas na revista "Estadística", órgão oficial de informação, e mediante boletins e folhetos distribuídos regularmente a seus associados.

PESQUISA ESPECIAL DE BOVINOS

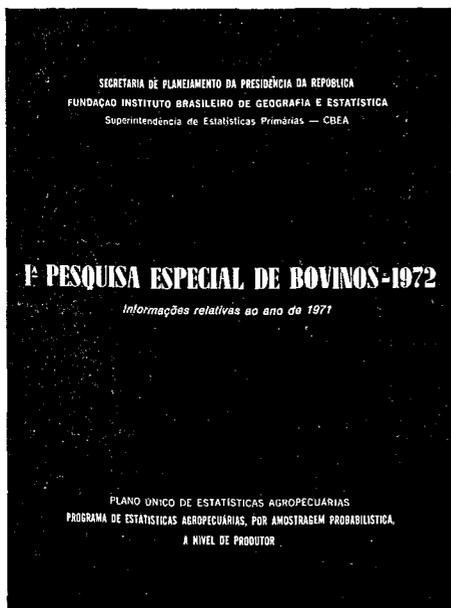
O IBGE acaba de divulgar os resultados da "1.^a PESQUISA ESPECIAL DE BOVINOS — 1972", elaborados com base em informações relativas ao ano de 1971, em oito Unidades da Federação: Bahia, Minas Gerais, São Paulo, Paraná, Santa Catarina, Rio Grande do Sul, Mato Grosso e Goiás, que no seu conjunto representavam 84,14% do rebanho bovino do País,

de acordo com os dados do Censo Agropecuário de 1970.

Referida Pesquisa faz parte do Programa de Estatísticas Agropecuárias por amostragem probabilísticas, a nível de produtor, que integra o Plano Único de Estatísticas Agropecuárias consideradas essenciais ao planejamento sócio-econômico do País e a segurança nacional.

A apresentação dos dados é efetuada por um conjunto de sessenta tabelas, sendo vinte e quatro por Grupos de Área dos estabelecimentos agropecuários, vinte e quatro por Grupos de Cabeças de Bovinos, e doze por Grupamentos Especiais, sendo que este último relaciona: Estabelecimentos por Grupos de Cabeças de Bovinos segundo Grupo de Área; Composição do rebanho segundo Fonte Principal de Renda; Produção de leite e outras variáveis segundo Fonte Principal de Renda; e Estabelecimentos com ordenha segundo Grupos de Cabeças de Vacas Ordenhadas.

Trata-se, pois, de trabalho pioneiro de grande profundidade, rea-



lizado dentro da mais moderna técnica da ciência estatística, que focaliza um dos mais importantes setores da economia brasileira, e

cujos resultados bem demonstram a grandiosidade da Pesquisa, bem como a evolução altamente positiva da agropecuária brasileira.

CENSO DOS SERVIÇOS

Estado do Amazonas

O IBGE inicia, com a apresentação do volume correspondente ao Estado do Amazonas, a divulgação dos dados definitivos do CENSO DOS SERVIÇOS, levantamento este que integrou o Recenseamento Geral de 1970, e cuja coleta de informações foi iniciada no segundo semestre de 1971.

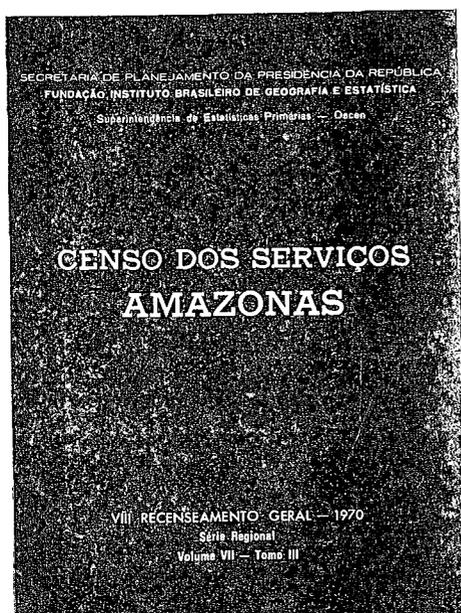
O Censo em questão, como os demais, revestiu-se de todas as características técnicas de um empreendimento dessa relevância, atendendo, inclusive, solicitações da Organização das Nações Unidas (ONU) e de outros organismos internacionais, integrando-se no Censo das Américas, patrocinado pelo Instituto Interamericano de

Estatística (IASI), órgão pertencente à Organização dos Estados Americanos (OEA), através da adoção do seu Programa Mínimo de investigações e tabulações, que visa a assegurar a uniformidade de conceitos e a comparabilidade dos resultados dos Censos das Nações Unidas.

As atividades abrangidas pelo Censo, são as que resultam da prestação de serviço, com fins lucrativos, a pessoas ou entidades, e foram classificadas em: Serviços de Alojamento e Alimentação; Serviços de Reparação, Manutenção e Conservação; Serviços Pessoais; Serviços Comerciais; e Serviços de Diversões, Radiodifusão e Televisão, e de Promoção de Espetáculos Artísticos.

O plano de divulgação compõe-se de 24 tomos relativos as Unidades da Federação (Série Regional), e um referente ao Brasil (Série Nacional).

A divulgação dos dados, no que diz respeito as características dos Estabelecimentos, se processa através de um conjunto de treze tabulações, sendo: uma de confronto dos resultados dos Censos de 1960 e 1970; uma de Dados Gerais — pessoal ocupado, média mensal do pessoal ocupado, despesas diversas, despesas de operações e valor da receita — a nível de Microrregiões



Homogêneas, Municípios e classes de serviços; e onze a nível de classes e gêneros de serviços, com resultados para o conjunto da Unidade da Federação.

O volume ora publicado se constitui num excelente repositório de dados, indispensáveis ao estudo econômico-social dos agrupamentos que enfeixa.

CENSO AGROPECUÁRIO

Paraná

O IBGE vem de publicar mais um volume do CENSO AGROPECUÁRIO, este relativo ao Estado do Paraná, contendo os resultados definitivos desse levantamento que integrou o Recenseamento Geral de 1970.

Tal publicação, permite-nos conhecer com segurança e detalhadamente todas as atividades agropecuárias dessa importante Unidade da Federação, cujos dados são apresentados através de 109 tabelas, das quais 57 a nível de Unidades da Federação e 52 a nível de Microrregiões Homogêneas e Municípios, abrangendo todos os aspectos investigados: estabelecimento, área total, propriedade das terras, condição legal das terras, condição do produtor, direção dos trabalhos, utilização das terras, terras irrigadas, atividade econômica, associação e cooperativas, pessoal ocupado, utilização do pessoal temporário, serviços de empreitada, fertilizantes, emprego de força, máquinas e instrumentos agrícolas, veículos, silos e depósito, valor dos bens, investimentos, financiamentos, despesas, combustíveis, energia elétrica,

ca, pecuária, aves, coelhos e colmeias, produção de origem animal, produção de origem vegetal, transformação e beneficiamento de produtos agropecuários, animais pertencentes ao pessoal residente nos estabelecimentos, produção particular do pessoal residente, etc.

Cada um dos itens citados comporta inúmeros detalhes que estão devidamente enumerados nas tabelas apresentadas.

A Pesquisa revestiu-se de mais rigorosa técnica e contou com a coordenação internacional (ONU, IASI e OEA), integrando-se no Censo das Américas, através da adoção do seu Programa Mínimo de investigações e tabulações, que visa a assegurar a uniformidade de conceitos e a comparabilidade dos resultados dos Censos das Nações Americanas.

Seus resultados permitem o estudo minucioso da vida agropecuária do Estado do Paraná e dos Municípios que o integram, permitindo, inclusive avaliar a sua evolução pelos dados comparativos de outros censos, que também apresenta.

BOLETIM INFORMATIVO

A SOCIEDADE BRASILEIRA DE ESTATÍSTICA, preenchendo uma lacuna que se fazia sentir nos seus meios de comunicação, vem de apresentar o primeiro número do "Boletim Informativo", datado de junho corrente, com o qual se propõe divulgar assuntos de interesse de seus associados, a colaboração destes, além de noticiário geral sobre atividades programadas ou não.

A iniciativa é oportuníssima e louvável, pois além do Boletim servir de veículo de comunicação com

os associados, estimulando o interesse associativo pela inserção de estudos de cunho didático e trabalhos técnicos de ressonância internacional, serve igualmente para melhor difundir as promoções culturais da entidade e, conseqüentemente, contribuir para o aprimoramento técnico do Estatístico.

Este primeiro número, noticioso e bem planejado, transmite a todos a certeza de que o periódico em questão, alcançará plenamente a sua finalidade.

PUBLICAÇÕES EDITADAS PELOS ÓRGÃOS DE ESTATÍSTICA DO IBGE NO TRIMESTRE ABRIL-JUNHO DE 1975 *

CENTRO DE DOCUMENTAÇÃO E INFORMAÇÃO ESTATÍSTICA — CENDIE

016:3(81)

BOLETIM BIBLIOGRÁFICO. Rio de Janeiro, v. 5, n. 3-4, jul./set. — out./dez. 1972 Trimestral.

025.346

BOLETIM TRIMESTRAL DA BIBLIOTECA WALDEMAR LO-

PES. Rio de Janeiro, v. 4, n. 3, jul./set. 1973.

CATALOGO DE PERIÓDICOS E PUBLICAÇÕES SERIADAS DA BIBLIOTECA WALDEMAR LOPES. Rio de Janeiro, v. 2, pt. 1, 1975.

* Bibliografia preparada pelo Setor de Bibliografia do Centro de Documentação e Informação Estatística do IBGE

DEPARTAMENTO DE DIVULGAÇÃO ESTATÍSTICA — DEDIVE

Periódicos

- 31(81) (05) jun. — jul./set. 1974. Trimestral.
BOLETIM ESTATÍSTICO. Rio de Janeiro, v. 32, n. 126-127, abr./

Séries

- 911.6(816.12 AGU LIN) de Janeiro, 1975. 24 p., il. (Coleção de Monografias, 587)
Águas de Lindóia, São Paulo. Texto de Maria de Lourdes F. Cianella. Rio de Janeiro, 1975. 24 p., il. (Coleção de Monografias, 583)
- 911.6(816.12 BAR BON) *Barra Bonita, São Paulo.* Texto de Umberto Cognac. Rio de Janeiro, 1975. 28 p., il. (Coleção de Monografias, 585)
- 911.6(816.12 CAM) *Campinas, São Paulo.* Texto de Aldalita de Jesus B. L. de Medeiros. Rio de Janeiro, 1975. 52 p., il. (Coleção de Monografias, 584)
- 911.6(816.12 DIA) *Diadema, São Paulo.* Texto de Maria de Lourdes F. Cianella. Rio de Janeiro, 1975. 24 p., il. (Coleção de Monografias, 589)
- 911.6(816.12 MIR) *Mirandópolis, São Paulo.* Texto de Aldalita de Jesus B. L. de Medeiros. Rio de Janeiro, 1975. 20 p., il. (Coleção de Monografias, 588)
- 911.6(816.12 OSA) *Osasco, São Paulo.* Texto de Aldalita de Jesus B. L. de Medeiros. Rio de Janeiro, 1975. 32 p., il. (Coleção de Monografias, 590)
- 911.6.(816. PAL MIS) *Palmeira das Missões, Rio Grande do Sul.* Texto de Aldalita de Jesus B. L. de Medeiros. Rio de Janeiro, 1975. 24 p., il. (Coleção de Monografias, 589)

CENTRO BRASILEIRO DE ESTUDOS DEMOGRÁFICOS — CBED

- 312(81) jan./mar. — abr./jun. 1975. Trimestral.
BOLETIM DEMOGRÁFICO CBED. Rio de Janeiro, v. 5, n. 3-4,

DEPARTAMENTO DE ESTATÍSTICAS INDUSTRIAIS, COMERCIAIS
E DE SERVIÇOS. — DEICOM

31:66/69(81)

Indústrias de transformação. Pesquisa mensal — janeiro/dezembro de 1974. Rio de Janeiro, 1975. 54 p., tab. Mimeografado.

— — *janeiro de 1975.* Rio de Janeiro, 1975. 22 p., tab. Mimeografado.

31:69(81)

Indústria da construção. Inquérito mensal sobre edificações — janeiro de 1975. Rio de Janeiro, 1975. 249 p., tab.

— — *fevereiro de 1975.* Rio de Janeiro, 1975. 243 p., tab.

338.5:31(81)

— — *Preços de material de construção no comércio atacadista. Salários na indústria da construção — outubro de 1974 a março de 1975.* Rio de Janeiro, 1975. 323 p., tab.

338.5:31(81)

Inquérito nacional de preços. Gêneros alimentícios e artigos do vestuário. Comércio atacadista e varejista nas Capitais — dezembro de 1974. Rio de Janeiro,

1975. 37 p., tab. Mimeografado.

— — *janeiro de 1975.* Rio de Janeiro, 1975. 37 p., tab. Mimeografado.

— — *fevereiro de 1975.* Rio de Janeiro, 1975. 37 p., tab. Mimeografado.

— — *março de 1975.* Rio de Janeiro, 1975. 37 p., tab. Mimeografado.

— — *abril de 1975.* Rio de Janeiro, 1975. 37 p., tab. Mimeografado.

— — *Comércio atacadista e varejista nas Unidades da Federação — dezembro de 1974.* Rio de Janeiro, 1975. 49 p., tab. Mimeografado.

— — *janeiro de 1975.* Rio de Janeiro, 1975. 49 p., tab. Mimeografado.

— — *fevereiro de 1975.* Rio de Janeiro, 1975. 49 p., tab. Mimeografado.

— — *março de 1975.* Rio de Janeiro, 1975. 49 p., tab. Mimeografado.

GRUPO EXECUTIVO DE PESQUISAS DOMICILIARES — GEPD

311.213.2:312.5(81)

Pesquisa Nacional por Amostra de Domicílios. População. Mão-de-Obra. Salário. Instrução.

Domicílio. 4. trimestre 1973. 2. tiragem. Rio de Janeiro, 1975. 383 p. mapa, tab. (Doc. GEPD, 61)

DEPARTAMENTO DE CENSOS — DECEN

62/69(81)

Pesquisa industrial — 1972 — Brasil. Aspectos gerais das atividades industriais. Rio de Janeiro, 1975. 120 p., tab.

311.213.1:63(812.2)

Censo Agropecuário — Piauí. Rio de Janeiro, 1975. XXXIX + 331 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 3, t. 6)

311.213.1:63(816.2)

— — *Paraná.* Rio de Janeiro, 1975. XLIII + 629 p., tab., questionário em anexo. (VIII Recenseamento Geral — 1970. Série Regional v. 3, t. 19)

311.213.1:63(817.2)

— — *Mato Grosso.* Rio de Janeiro, 1975. XXXIX + 357 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 3, t. 22)

311.213.1:658.64(811.1 + 811.4 + + 811.6)

Censo dos Serviços — Rondônia — Roraima — Amapá. Rio de Janeiro, 1975. XXIII + 160 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 1)

311.213.1:658.64(811.2)

— — *Acre.* Rio de Janeiro, 1975. XXIII + 53 p., tab. questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 2)

311.213.1:658.64(811.3)

— — *Amazonas.* Rio de Janeiro, 1975. XVI + 59 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 3)

311.213.1:658.64(811.5)

— — *Pará.* Rio de Janeiro, 1975. XXIV + 67 p., tab., questionário em anexo. (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 4)

311.213.1:658.64(812.1)

— — *Maranhão.* Rio de Janeiro, 1975. XXV + 75 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 5)

311.213.1:658.64(812.2)

— — *Piauí.* Rio de Janeiro, 1975. XXV + 71 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 6)

311.213.1:658.64(813.1)

— — *Ceará.* Rio de Janeiro 1975. XVI + 81 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 7)

311.213.1:658.64(813.2)

— — *Rio Grande do Norte.* Rio de Janeiro, 1975. XXVI + 79 p., tab. questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 8)

311.213.1:658.64(813.3)

— — *Paraíba*. Rio de Janeiro, 1975. XXVI + 83 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 9)

311.213.1:658.64(813.4)

— — *Pernambuco*. Rio de Janeiro, 1975. XXVII + 87 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 10)

311.213.1:658.64(813.5)

— — *Alagoas*. Rio de Janeiro, 1975. XXV + 71 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 11)

311.213.1:658.64(814.1)

— — *Sergipe*. Rio de Janeiro, 1975. XXIV + 67 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 12)

311.213.1:658.64(814.2)

— — *Bahia*. Rio de Janeiro, 1975. XXX + 117 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 13)

311.213.1:658.64(815.2)

— — *Espírito Santo*. Rio de Janeiro, 1975. XXVI + 63 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 15)

311.213.1:658.64(815.3)

— — *Rio de Janeiro*. Rio de Janeiro, 1975. XXIV + 67 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 16)

311.213.1:658.64(816.2)

— — *Paraná*. Rio de Janeiro, 1975. XXIX + 115 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 19)

311.213.1:658.64(816.4)

— — *Santa Catarina*. Rio de Janeiro, 1975. XXVII + 93 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 20)

311.213.1:658.64(817.2)

— — *Mato Grosso*. Rio de Janeiro, 1975. XXVI + 69 p., tab., questionário em anexo (VIII Recenseamento Geral 1970. Série Regional v. 7 t. 22)

311.213.1:658.64(817.3)

— — *Goiás*. Rio de Janeiro, 1975. XXX + 95 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 23)

311.213.1:658.64(817.4)

— — *Distrito Federal*. Rio de Janeiro, 1975. XVI + 51 p., tab., questionário em anexo (VIII Recenseamento Geral — 1970. Série Regional v. 7, t. 24)

CENSO DOS SERVIÇOS

Rondônia, Roraima, Amapá, Maranhão, Ceará, Rio Grande do Norte, Alagoas, Sergipe, Espírito Santo, Santa Catarina e Distrito Federal

Dando continuidade a divulgação dos resultados definitivos do CENSO DOS SERVIÇOS, iniciada com o volume correspondente ao Estado do Amazonas, o IBGE, vem de distribuir mais os volumes correspondentes aos Territórios de RONDÔNIA, RORAIMA e AMAPÁ, aos Estados do MARANHÃO, PIAUÍ, CEARÁ, RIO GRANDE DO NORTE, PARAÍBA, ALAGOAS, SERGIPE, ESPÍRITO SANTO e SANTA CATARINA, e também, ao DISTRITO FEDERAL.

Referido Censo, que integrou o Recenseamento Geral de 1970, foi processado dentro do maior rigorismo técnico, obedecendo normas internacionais sobre a matéria, inclusive quanto ao atendimento de recomendações de organismos internacionais (ONU, OEA e IASI), objetivando a uniformidade de conceitos e a comparabilidade dos Censos das Nações Unidas.

A divulgação dos dados é feita através de um conjunto de treze tabulações, sendo: uma de Dados Gerais — pessoal ocupado, média mensal do pessoal ocupado, salários, despesas diversas, despesas de operação e valor da receita — a nível de Microrregiões Homogêneas, Municípios e classes de serviços; e onze a nível de classes e gêneros de serviços. As atividades de Prestação de serviços foram classificadas em: Serviços de Alojamento e Alimentação; Serviços de Reparação, Manutenção e Conservação; Serviços Pessoais; Serviços Comerciais; e Serviços de Diversões, Radiodifusão e Televisão, e de Promoção de Espetáculos Artísticos.

Os dados ora apresentados oferecem aos estudiosos da ciência estatística, e órgãos de planejamento, elementos dos mais valiosos para a análise e avaliação do campo sócio-econômico em que as respectivas atividades se enquadram.

Todas as publicações do IBGE, inclusive as referentes as notas insertas nesta Revista, poderão ser adquiridas na sede da Instituição, à Avenida Franklin Roosevelt, 146 — loja, Rio de Janeiro — RJ, ou nas respectivas Delegacias de Estatística, nas demais Unidades da Federação.

Composto e impresso no
Centro de Serviços Gráficos
do IBGE, Rio de Janeiro, RJ

IBGE

Presidente: ISAAC KERSTENETZKY

Diretor-Geral: EURICO DE ANDRADE NEVES BORBA

Diretor-Técnico: AMARO DA COSTA MONTEIRO

DEPARTAMENTO DE DIVULGAÇÃO ESTATÍSTICA

Chefe-Substituto: Márlo Fernandes Paulo

SECRETARIA DA REVISTA BRASILEIRA DE ESTATÍSTICA

Chefe: Fernando Pereira Cardim