

Análise fatorial: problemas e aplicações na geografia, especialmente nos estudos urbanos

SPERIDIÃO FAISSOL
Geógrafo do IBG

1. Introdução

○ propósito do presente trabalho é o de apresentar alguns dos problemas conceituais básicos e implicações teóricas do uso da análise fatorial, particularmente aplicável ao caso de estudos urbanos. Por isso mesmo, ao lado de proposições que são essencialmente de natureza conceitual filosófica e que são premissas intrínsecas ao uso da metodologia fatorial, colocamos problemas teóricos que ficam também implícitos no contexto de decisões arbitrárias, que são necessariamente tomadas no processo de pesquisa.

Para evitar que tais problemas e proposições fiquem no terreno das abstrações, uma série de análises, feitas do sistema urbano brasileiro, é utilizada, ao mesmo tempo, como guia metodológico (embora simplificado) e como exemplificações da adoção de premissas diferentes ou de decisões arbitrárias, como foi indicado acima. As análises feitas foram em número de seis, duas para o sistema nacional, uma usando variáveis absolutas ou relativas não transformadas e outra realizando uma transformação logarítmica prévia dos dados, tanto os absolutos como os relativos; as quatro outras foram feitas obedecendo-se ao mesmo mecanismo (variáveis simples e transformadas), uma para a área subdesenvolvida e outra para a área desenvolvida (mais ou menos o Núcleo e a Periferia).

É claro que os exemplos apresentados, na realidade as próprias decisões tomadas no contexto do processo de análise, já refletem hipóteses teóricas sobre o sistema urbano brasileiro, já apresentadas em algumas diferentes oportunidades pelo autor (1), o que na realidade significa que o trabalho pretende contribuir menos a um melhor entendimento do sistema urbano do que a uma análise dos problemas metodológicos e implicações teóricas de um estudo de ecologia fatorial.

COMENTÁRIOS

2. O significado da análise fatorial como metodologia analítica e suas implicações teóricas

Duas decisões arbitrárias fundamentais são tomadas no processo da ecologia fatorial:

a) A escolha das unidades observacionais — escolha que implica, ao mesmo tempo, no número de unidades representativas de uma coletividade qualquer e o seu grau de agregação a serem incluídas na análise.

b) A escolha das variáveis subscritas — escolha que também implica numa gama de aspectos que sejam teoricamente pertinentes e níveis de agregação destas variáveis compatíveis com os objetivos da pesquisa que devem ser incluídas na análise.

Subjacentes a estas duas decisões arbitrárias estão dois problemas igualmente fundamentais em qualquer tipo de análise, representados pelas diferentes falácias (das quais as duas mais importantes são a individualista e a ecológica) e pelas implicações teóricas do uso do método propriamente dito. Ambas serão tratadas no contexto das duas outras acima indicadas, ilustradas com exemplos das análises feitas e referidas na introdução.

A chamada falácia ecológica foi primeiro proposta por W. S. Robinson, que procurou demonstrar e “provar” que associações estatísticas para populações agregadas podem diferir em magnitude e mesmo em sinal das associações para os indivíduos que compõem aquela população agregada (2), através de agregações diferentes Robinson verificou que a nível do que ele chamou de indivíduos, em uma população, nos Estados Unidos, observava-se uma correlação muito baixa entre analfabetismo e negros, mas que tal correlação podia ser elevada “por efeitos ecológicos espúrios a uma fantástica magnitude de 0,95, quando dados agregados para um pequeno número de regiões são analisados.” Esta citação é indireta, provinda do artigo de Alker na coleção de leituras editada por Dogan e Rokkan que, logo a seguir, procuram mostrar a direção reversa do raciocínio de Robinson, e que representa a chamada falácia individualista, ao se perguntar se as pessoas sempre confundem relações entre indivíduos com as relações entre agregados; certamente que sim, diz Alker, quando cientistas sociais ideologicamente motivados tendem a generalizar, do indivíduo para o coletivo. Devemos acreditar, pergunta Alker, que “por que os trabalhadores são socialmente radicais, haverá uma elevada tendência semelhante para exprimir opiniões radicais nos países altamente desenvolvidos com alta proporção de trabalhadores em suas populações? Ou os trabalhadores nestes países tendem a ser mais conservadores?” pp. 78 (3).

2.a) A escolha das unidades observacionais

Começaremos por analisar o problema das unidades observacionais adotadas que contém, em seu bojo, o problema das duas falácias acima mencionadas. Preliminarmente é claro que a questão essencial relativa à unidade observacional é a do objetivo do estudo e vamos, daqui por diante, restringir o âmbito da discussão àquilo que ela circunscreve especificamente, quer dizer, o estudo de um sistema de cidades. Isto torna apenas aparentemente mais fácil a análise da questão unidade

observacional, uma vez que a cidade seria a unidade. Entretanto, mesmo aí, existem problemas dos mais sérios, teórico, conceituais e mesmo práticos.

Considerando um ângulo analítico sistêmico, existe o problema de definir o sistema, isto é, enumerar as cidades que o constituiriam. Pareceria fácil fazer uma escolha arbitrária — vamos dizer as cidades de mais de 10 mil habitantes do Brasil.

Mas qual a justificativa para isso? As cidades de dez mil habitantes são aquelas que têm um mínimo de serviços urbanos e atividades não agrícolas, com um mínimo de diversificação, capaz de diferenciar o agrupamento populacional de outros, que por sua natureza sejam essencialmente rurais? Um critério deste tipo é válido para todo o Brasil? O que isto quer dizer é que podemos considerar uma cidade de dez mil habitantes no Rio Grande do Sul, em São Paulo, no Nordeste e na Amazônia, como limite aceitável entre um núcleo efetivamente urbano, com um modo diferente de viver do rural, em que o indivíduo que para ele migrou adotou uma forma urbana de vida, rompendo com a tradição rural anterior? Este é um aspecto do problema. Outro aspecto igualmente importante é que cidades grandes constituem o ponto de convergência de um amplo mercado de trabalho, com um espaço em torno, organizado em termos de viagens diárias do domicílio para o trabalho, estabelecendo relações de unidades politicamente consideradas como cidades, às vezes de menos de 10 mil habitantes e às vezes tão grande como Nova Iguaçu, com centenas de milhares de habitantes. Elas devem ser tomadas como unidades observacionais diferentes, ou tomadas em conjunto, considerando uma área metropolitana, assim definida, como uma unidade observacional? Mesmo que assim fosse, e no caso de cidades menores muito associadas, do tipo Barra Mansa-Volta Redonda, como considerar o problema, sem adotar-se para cada caso uma decisão arbitrária, consistente ou não com uma realidade objetiva?

Poderíamos exemplificar, desde logo, utilizando dois resultados diferentes em duas análises feitas do sistema de cidades brasileiras (4), em que na primeira a unidade observacional foi a área metropolitana (para o caso das nove áreas) e numerosos agregados de cidades; na segunda, especificamente para as áreas metropolitanas tomou-se apenas as cidades centrais e alguns dos núcleos satélites destas áreas. No caso específico das áreas de São Paulo, verificou-se uma diferença substancial de tamanho funcional entre São Paulo e Rio, nas duas análises. No que diz respeito ao *status* econômico ou nível de urbanização, na primeira análise (agregada), uma sensível diferença se verificava entre os valores dos *scores* de São Paulo e Rio, o que na segunda análise (desagregada) não mais aparecia. Entretanto podia-se observar que as diferenças agregadas se deviam aos núcleos da área de S. Paulo e Rio, equivalentes às diferenças macrorregionais entre o Núcleo e a Periferia. O objetivo da desagregação havia sido explorado na análise, o de testar a hipótese de que haveria diferenças substanciais de grau de urbanização e equipamento urbano, *status* econômico das populações, etc., entre as duas áreas, no que se referia às suas áreas de expansão. Entretanto o caso serve ao mesmo propósito aqui indicado, e que é o de mostrar que diferentes níveis de agregação para as unidades observacionais podem conduzir a resultados finais que obscurecem a natureza do problema essencial que se procura pesquisar. Corroborar, inteiramente, a falácia ecológica, quase que ao mesmo tempo que corroboraria a falácia individualista, se alguém tentasse concluir,

pela análise apenas das duas cidades centrais, que não haveria diferença quase nenhuma entre as duas no seu nível de desenvolvimento, urbanização, etc.

Algumas destas questões geram outras igualmente importantes: Estamos considerando, dado o tema — importância da análise fatorial em estudo de um sistema urbano — e fazemos a indagação do limite que devemos estabelecer ao sistema e vamos, apenas para ilustrar, aceitar que tomamos as cidades de dez mil habitantes em 1960, ou mais precisamente, tomamos 209 cidades no Brasil como representativas do sistema. Embora o problema volte a ser discutido a seguir, vamos aceitar que 59 variáveis, como as utilizadas no trabalho que já citamos, sejam representativas do processo pelo qual as cidades se diferenciam umas das outras no Brasil, segundo as suas principais dimensões.

A primeira etapa do sistema de operações matriciais que a análise fatorial requer é o cálculo das médias e do desvio padrão em torno destas médias, para ser utilizado, posteriormente, nos cálculos dos *scores* das cidades nos diferentes fatores. Mas estas médias e seus desvios padrão já são, em si, representativas do desempenho de cada variável na formação dos fatores e muito mais que isso, na interpretação dos resultados. Esta média é obtida da forma clássica, quer dizer, a soma dos valores dividida pelo número de observações; uma variável que contenha um número elevado de valores zero vai ter uma média fortemente distorcida, um elevado desvio padrão e precisa ser escolhida somente se a espécie de poder discriminatório que ela vai significar tenha sentido teórico. Um exemplo é fácil de ser entendido e refere-se a variável número de alunos matriculados no curso superior por mil habitantes (portanto uma variável relativa), que obviamente é altamente discriminatória no sistema de cidades, na realidade no contexto da estrutura socioeconômica do Brasil. Mesmo considerando a transformação logarítmica da variável, a sua média nacional é 1,29, com um desvio padrão de 1,31. Comparada com a variável ensino secundário, observa-se que a média é de 3,62 e o desvio padrão 0,61, portanto, com uma universalidade muito maior que a primeira, óbvia para quem tem qualquer noção do processo educacional brasileiro. Quando subdividimos o sistema de cidades entre as 57 da periferia e as 152 do núcleo (aí tomado como o Centro-Sul, de um modo geral), verificamos que para a periferia a média do ensino secundário (mesma transformação logarítmica) é de 3,18 e o desvio padrão é de 0,80, embora para o núcleo a média seja de 3,78 e o desvio padrão seja de 0,43. No caso do ensino superior a média para o núcleo é de 1,50 e o desvio padrão é de 1,35, ao passo que na periferia a média é de 0,73 e o desvio padrão é de 1,01. Como no conjunto da análise obtém-se uma estrutura de relações, estes detalhes passam despercebidos em muitos casos, mas são relevantes. Em primeiro lugar mostra que para o ensino secundário os desvios da média nacional são pequenos, mas são um pouco maior na periferia que na região desenvolvida. Mas no caso do ensino superior, embora em termos nacionais, o desvio em relação à média seja superior ao valor da média, (1,29 e 1,31), a diferença é mínima; no que diz respeito à área subdesenvolvida o desvio é bastante grande (1,01 para a média de 0,73), enquanto que para o sistema de 152 cidades o desvio não está longe da média nacional (1,35 para média 1,50), com um coeficiente de variação bem menor que o da periferia e praticamente igual ao nacional.

Estas considerações podem levar à indagação de validade da variável no contexto nacional e não validade no contexto das 57 cidades (por outro lado, assim se perderia a comparabilidade). Considere-

rando que o método analítico parte da premissa da linearidade entre os dados, isto constituiria, a rigor, uma violação desta premissa; por outro lado, como escolher? O ensino superior é uma parcela significativa do processo de desenvolvimento brasileiro e por via desta conclusão indispensável no contexto de uma análise do sistema de cidades. Ou será que os contrastes do sistema urbano brasileiro são de tal natureza que seria falso defini-lo como um sistema nacional articulado (como um sistema precisa ser), e então teríamos apenas subsistemas, num dos quais a variável ensino superior poderia ser utilizada sem violação das premissas do método analítico e no outro não poderia? A tabela 1 mostra um número selecionado de variáveis, computadas as suas médias e desvios padrão, tanto para o plano nacional como para os dois sistemas referidos e muitas diferenciações semelhantes podem ser detectadas. Uma delas é a comparação entre número de prédios servidos por água e por esgoto (também variáveis relativas), que se diferenciam talvez ainda mais que as referentes a ensino secundário e superior. No entanto água e esgoto são serviços urbanos essenciais e a análise de um sistema urbano não pode prescindir das mesmas para identificar o nível de urbanização. Ao mesmo tempo estas duas variáveis são excelentes indicadores das sucessivas etapas do processo de urbanização. Ainda mais, pode-se considerar que o simples fato de que os desvios padrão, tanto delas (principalmente esgoto), são muito maiores no sistema de 57 cidades que no de 152, são indícios significativos de que os maiores contrastes de renda se localizam na área subdesenvolvida (inclusive porque isso ocorre com todos os indicadores), desde água ou esgoto, ensino superior e secundário, até automóveis ou pessoal ocupado na indústria pesada. Nesta última note-se que o peso do sistema de 152 cidades é tão grande que a média nacional é 2,51 e o desvio padrão é de 1,60 (inferior à média), enquanto que nas 57 cidades é de 0,69 e 0,93, portanto bem maior que a média, ao passo que no sistema de 152 cidades é de 1,80 e 1,26, muito menor que a média.

Até aqui estamos discutindo o comportamento de variáveis isoladamente em três contextos diferentes, um nacional e os dois outros regionais, em termos de uma concepção centro-periferia do modelo urbano brasileiro. Mas a análise fatorial visa, essencialmente, reduzir um número elevado de variáveis a um número menor de fatores, verdadeiras dimensões básicas do sistema, segundo as quais cada uma das cidades seria posicionada e, por fim, agrupadas em uma tipologia ou regionalização que descreveria o sistema. Como se comportam os fatores nestes três planos? São os mesmos e contém a mesma estrutura de relações, tanto no plano nacional como nos dois regionais mencionados? Esta é uma questão mais essencial ainda e nela está embutida a questão da chamada "falácia ecológica". Senão vejamos: Quando usamos as 209 cidades e as 59 variáveis, a análise extrai 11 fatores que em seu conjunto explicam quase 80% da variação, o primeiro referente ao tamanho funcional explicando 22,53, o segundo o nível de urbanização e desenvolvimento explicando mais 16,64 e o terceiro com mais 9,40 (total pouco inferior a 50%), sendo o terceiro referente à estrutura bipolarizada indústria/comércio e serviços. Não discutiremos aqui nem a propriedade de uma hipótese tamanho funcional nem a propriedade das variáveis que a definem, tomando por enquanto tal coisa como dada. Mas vejamos como se compõe o fator tamanho funcional no plano nacional e nos dois planos regionais, em termos da estrutura de relações interváveis. As variáveis que refletem o tamanho absoluto dos lugares — população, número de automóveis, pessoal ocupado na indústria,

TABELA 1

Coeficiente de variação para 15 variáveis

VARIÁVEIS	209 CIDADES C/TRANSE. LOGARÍTMICA			152 CIDADES C/TRANSE. LOGARÍTMICA			S/TRANSE. LOGARÍTMICA		
	x		ev.	x		ev.	x		ev.
Pop. total — 1967.....	0,69	0,66	96	0,67	0,66	99	50,95	204,82	402
Pes. ocupado na ind. — 1965.....	0,31	0,53	171	0,35	0,58	166	59,70	360,92	603
N.º de automóveis — 1967.....	0,35	0,55	157	10,21	310,60	146	61,22	345,86	535
Automóveis/1.000 hab. — 1967.....	2,80	0,99	35	3,22	0,66	20	23,16	14,49	63
N.º de alunos do curso su./1.000 hab. — 1967	1,29	1,31	102	1,50	1,35	90	5,37	7,82	146
N.º de prédios com água/1.000 hab. — 1967	3,19	0,94	29	3,43	0,67	19	104,79	51,58	49
N.º de prédios com esgotos/1.000 hab. — 1967	2,61	1,60	64	3,06	1,29	42	71,45	53,97	71
% do pess. ocup. na ind. pesada/pessoal ocup. na ind. total.....	1,56	1,29	83	1,89	1,26	67	12,13	17,10	141
% do pess. ocup. na ind. trad./total do pess. ocup. na ind. — 1965.....	3,66	0,79	21	3,63	0,82	23	45,55	22,38	49
% de pop. de 0 a 14 anos/pop. total.....	4,26	0,32	8	4,25	0,36	8	41,51	13,70	9
Pop. urbana (total)/ocup. na ind. em 1960 por 10.000 hab.....	2,56	0,88	31	2,68	0,85	32	510,40	441,33	86
Pop. urbana (total)/ocup. no comércio em 1960/10.000 hab.....	3,59	0,48	13	3,63	0,50	14	269,84	87,99	33
Pop. urbana (total)/ocupada em 10.000 hab. 1960.....	2,65	0,52	20	2,77	0,47	17	181,64	95,18	52
Distância da cidade central p/S. Paulo.....	3,15	0,90	29	2,74	0,68	25	9,77	5,87	60
Distância da cidade central p/Recife.....	2,73	0,80	29	3,11	0,24	8	44,16	9,68	22

TABELA 1

Coeficiente de variação para 15 variáveis das 57 cidades subdesenvolvidas

VARIÁVEIS	TRANSE. LOGARÍTMICA			S/TRANSE. LOGARÍTMICA		
	x		ev.	x		ev.
Pop. total — 1967.....	0,74	0,68	91	40,70	69,98	171
Pes. ocupado na ind. — 1965.....	0,23	0,33	143	16,23	35,51	218
N.º de automóveis — 1967.....	0,19	0,36	189	12,19	31,94	262
Automóveis/1.000 hab. — 1967.....	1,66	0,81	48	4,96	4,95	99
N.º de alunos do curso su./1.000 hab. — 1967	0,73	1,01	138	1,56	2,49	159
N.º de prédios com água/1.000 hab. — 1967	2,41	1,12	46	45,02	34,10	75
N.º de prédios com esgotos/1.000 hab. — 1967	1,02	1,40	137	14,88	26,40	177
% do pess. ocup. na ind. pesada/pessoal ocup. na ind. total.....	0,69	0,93	134	2,30	4,34	188
% do pess. ocup. na ind. trad./total do pess. ocup. na ind. — 1965.....	3,87	0,68	17	54,84	23,96	43
% de pop. de 0 a 14 anos/pop. total.....	4,29	0,11	2	42,00	3,47	8
Pop. urbana (total)/ocup. na ind. em 1960 por 10.000 hab.....	2,21	0,86	38	314,77	423,67	122
Pop. urbana (total)/ocup. no comércio em 1960/10.000 hab.....	3,48	0,38	10	230,75	75,44	32
Pop. urbana (total)/ocupada em 10.000 hab. 1960.....	2,33	0,52	22	117,51	55,25	47
Distância da cidade central p/S. Paulo.....	4,24	0,35	8	39,82	9,37	23
Distância da cidade central p/Recife.....	1,70	0,85	50	14,05	15,37	109

no comércio e nos serviços, etc. Naturalmente correlacionam-se fortemente em tal fator, tanto na análise a nível nacional como nas duas análises de partes do sistema nacional; o que difere nas duas análises regionais é que na relativa ao sistema de 57 cidades, além de terem sido extraídos 14 fatores em vez de 11 na de 152 cidades, variáveis relativas ao nível de urbanização e desenvolvimento aparecem correlacionadas, de forma acentuada, conforme se pode observar na tabela 2, que mostra a composição estrutural do fator I (Tamanho Funcional) nas três análises. Por esta tabela verifica-se que o tamanho funcional está associado ao nível de desenvolvimento (automóveis, telefones, médicos, leitos em hospitais, profissões liberais), à presença de maior quantidade de escolas superiores (e em conseqüência maior número de alunos matriculados) e à existência de uma indústria pesada (obviamente quase que associada aos grandes núcleos metropolitanos regionais). O significado desta associação parece claramente ligado ao fato de que os mecanismos de difusão funcionam pouco nas áreas subdesenvolvidas, e os mais elevados níveis de progresso se concentram nas grandes cidades. A grande cidade é grande e mais desenvolvida que as pequenas, o que pode ser facilmente identificado nos valores dos *scores* de metrópoles regionais e capitais de estado da área, constantes da tabela 3 e referentes aos fatores I e IV que definem o tamanho funcional e uma dimensão socioeconômica, um misto de *status* social e econômico.

Já no sistema de 152 cidades, embora o número de médicos e outras profissões liberais apareçam correlacionadas com o tamanho funcional, nenhuma das variáveis de indicadores de desenvolvimento aparece em tal fator, tanto que metrópoles do tipo de Belo Horizonte e Porto Alegre não têm *scores* muito diferentes de centros regionais menores do tipo de Campinas, Ribeirão Preto ou mesmo Piracicaba. Isto quer dizer que há uma dimensão tamanho e outra nível de desenvolvimento, efetivamente independentes uma da outra, diferentemente do que ocorre no sistema de 57 cidades. O importante na comparação é que na análise ao nível das 209 cidades, o peso das 152 e de seus valores mais elevados obscurece a realidade de que esta não é uma dimensão nacional e sim regional, característica do núcleo básico brasileiro; quanto à periferia, ela segue o padrão clássico de associação de tamanho e desenvolvimento, ligado a uma etapa do processo de desenvolvimento.

Estas comparações procuram, à luz de exemplos brasileiros conhecidos, apenas mostrar o elevado significado da escolha adequada e definição precisa das unidades observacionais, seja para agregação das mesmas em unidades que, por assim dizer, façam sentido, seja para definir o conjunto e subconjuntos de cidades que realmente representem o processo de urbanização nas suas múltiplas facetas e nas suas relações com o processo de desenvolvimento econômico. Entretanto isso parece apenas colocar o problema como de difícil solução, quem sabe sem solução, sem discutir seu fundamento teórico conceitual e suas implicações filosóficas.

Poderíamos iniciar a discussão deste aspecto do problema, citando uma síntese final da parte II (*Individual behaviour and collective properties*) do livro editado por Mattei Dogan e Stein Rokkan: "Em suma, esta seção de nosso volume defende o confronto de levantamentos ecológicos com levantamentos individuais, análises regionais e nacionais. A polarização levantamento individual/ecológico parece ser profundamente falsa. O que é necessário é uma estratégia de ligação sistemática: um esforço deliberado para fundir informações em vários níveis de variação" pp. 89 (5).

Na realidade o que esta síntese procura evidenciar é o que BERRY chama de “dialética entre o modelo e realidade”, a *interface* entre a chamada falácia ecológica e a falácia individualista, só resolvida com uma concepção sistêmica, na qual estruturas de inter-relações são válidas apenas no contexto em que elas são formuladas e que, para serem bem formuladas, precisam partir de uma concepção ou modelo da realidade, e as evidências empíricas que acumulamos são usadas para testar o modelo. Uma série de evidências repetitivas acabam por produzir suficiente informação e induzem à formulação de teoria explicativa. A dificuldade reside no fato de que concebemos o modelo e o usamos no próprio ato de coletar evidências empíricas para validar o modelo e só um esforço deliberado de confrontar resultados, uma espécie de tentativa de recuar dos conceitos e hipóteses que construímos, quase que ao longo de nossa formação profissional, para tentar ver a realidade “como ela realmente é” e testar o modelo. Tal dialética é tão infundável como a tentativa de descobrir a verdade das coisas.

A análise quantitativa é um poderoso ingrediente nesta dialética. Em primeiro lugar porque com os modernos computadores digitais de grande porte, a capacidade analítica das ciências sociais está altamente ampliada, tanto em termos dos aspectos da realidade que pode estudar, como pela poderosa bateria de algoritmos organizados que existem para testar hipóteses, descrever estruturas de relações, similar situações do mundo real, etc.

Em segundo lugar porque, embora a linguagem matemática não acrescente conteúdo à linguagem de qualquer ramo das ciências sociais ou biológicas, pois a sua própria essência é ser abstrata, ela contém

TABELA 2

Variáveis que têm correlação superior a 0,40 no fator I nas 3 análises com transformação logarítmica

VARIÁVEIS	209	152	57
1 — Pop. total — 1967.....	0,97	0,98	0,97
2 — Pes. ocup. ind. — 1965.....	0,90	0,91	0,92
3 — Pes. ocup. ind. — 1960.....	0,90	0,91	0,91
4 — Pes. ocup. com. — 1960.....	0,98	0,98	0,98
5 — Pes. ocup. serv. — 1960.....	0,96	0,96	0,97
6 — N.º estab. atacad. — 1960.....	0,96	0,96	0,91
7 — N.º estab. varej. — 1960.....	0,95	0,97	0,95
8 — N.º est. ind. — 1960.....	0,92	0,93	0,93
9 — N.º estab. serv. — 1960.....	0,95	0,96	0,96
10 — N.º estab. ensino MD — 1967.....	0,92	0,92	0,92
11 — N.º de automóveis — 1967.....	0,93	0,96	0,95
13 — N.º de Instituições de difusão/100.000 hab.....	—	—0,40	—
19 — Receita de serviços/pes. ocup. serv. — 1960.....	—	0,40	—
20 — Automóveis/1.000 hab. — 1967.....	—	—	0,63
21 — Telefones/1.000 hab. — 1967.....	—	—	0,42
23 — N.º médicos/10.000 hab. — 1967.....	0,42	0,49	0,49
25 — N.º dent. e farmacêuticos/10.000 hab. 1967.....	—	—	0,48
26 — N.º de outras profissões liberais/10.000 hab. — 1967	0,51	0,56	0,66
30 — N.º alunos curso superior/1.000 hab. 1967.....	—	—	0,76
36 — % do pes. ocup. ind. pesada/pes. ocupado ind. total — 1965.....	—	—	0,47
49 — Densidade pop. — 1967.....	0,54	0,53	0,53
42 — % do maior setor industrial em valor vendas.....	—0,46	—0,45	—
50 — pop. da área de influência (\div 1.000) — 1967.....	0,83	0,84	0,87
58 — N.º de leitos — 1967.....	0,88	0,88	0,93

TABELA 3

CIDADES	TAMANHO		DESENVOLVIMENTO	
	Fator I		Fator IV	Fator II
	57	152	57	152
Belém.....	32,2874	---	12,0158	---
Belo Horizonte.....	---	46,1418	---	11,9472
Curitiba.....	---	37,2506	---	7,6972
Fortaleza.....	36,8495	---	8,5885	---
Porto Alegre.....	---	51,8259	---	12,6207
Recife.....	56,9848	---	11,3351	---
Rio de Janeiro.....	---	80,2297	---	16,1173
Salvador.....	47,5379	---	13,0744	---
São Paulo.....	---	87,8555	---	17,6743
Aracaju.....	17,0428	---	11,6848	---
Cuiabá.....	---	3,2602	---	2,4456
Florianópolis.....	---	8,7959	---	7,9416
Goânia.....	---	14,6532	---	4,6468
João Pessoa.....	20,3913	---	10,5343	---
Maceió.....	21,7539	---	11,4390	---
Manaus.....	20,9672	---	11,6858	---
Natal.....	19,8536	---	13,3009	---
São Luís.....	20,8141	---	10,5927	---
Vitória.....	---	16,4921	---	3,2176
Terezina.....	8,7237	---	6,5380	---
Macapá.....	--- 4,5102	---	1,7669	---
Porto Velho.....	---10,5157	---	--- 3,9621	---
Rio Branco.....	--- 7,9311	---	--- 0,5654	---
Campina Grande.....	17,9143	---	10,3966	---
Campinas.....	---	30,6594	---	18,8491
Ribeirão Preto.....	---	18,7554	---	14,1915
Caxias do Sul.....	---	10,0967	---	8,3011
Joinville.....	---	6,0959	---	5,0461
São José dos Campos.....	---	3,9390	---	9,3477
Jundiá.....	---	11,8571	---	8,9788
Pelotas.....	---	20,9809	---	5,4468
Piracicaba.....	---	11,8521	---	11,2164
Santos.....	---	35,7515	---	15,3350
São José do Rio Preto.....	---	9,3574	---	13,1194
Sorocaba.....	---	13,6395	---	10,1438
Taubaté.....	---	8,2409	---	8,3156
Uberaba.....	---	4,8606	---	6,0988
Uberlândia.....	---	6,2343	---	5,1683
Volta Redonda.....	---	8,2139	---	4,6909
Blumenau.....	---	7,8660	---	6,5433
Limeira.....	---	3,0997	---	6,5850
Rio Claro.....	---	5,0083	---	9,7257
São Carlos.....	---	4,8302	---	10,7951
Poços de Caldas.....	---	0,3171	---	6,6339
Campo Grande.....	---	5,8093	---	3,0964
Londrina.....	---	9,8542	---	0,8464
Garanhuns.....	--- 2,6151	---	2,2050	---
Santarém.....	--- 6,7359	---	--- 5,7703	---
Montes Claros.....	---	--- 0,7047	---	--- 7,6601
Apucarana.....	---	4,5046	---	8,5078
Maringá.....	---	3,0550	---	--- 5,2931

no seu bojo o fundamento do argumento lógico. Ela obriga como diz HARVEY (6) a pensarmos logicamente. KARL DEUTSCH (7) inicia a primeira parte do livro editado por DOGAN e ROKKAN (8) *The logic of Ecological Inference*, com uma descrição das três maiores funções dos métodos quantitativos e que são elaboração de teorias existentes e sua

clara especificação, a geração de sugestões heurísticas sobre fenômenos novos e por fim o uso de métodos quantitativos para decidir questões teóricas, através de processos do que ele chama forte inferência.

Como se vê, a questão fundamental é teorização, formulação de hipóteses que possam ser testadas e por via de seqüências repetitivas de regularidades que se possam observar em fenômenos, formando padrões que se repetem em diferentes testes, formular teorias. Em outras palavras teorização *a priori* ou *a posteriori* constitui o cerne do processo de pesquisa, na qual os métodos quantitativos de inferência constituem, hoje em dia, instrumentos de um poder de que nunca se dispôs anteriormente nas ciências sociais.

Em suma, o único guia indesejável do processo de seleção de lugares é o da formulação teórica do problema a ser analisado e o seu teste através das técnicas apropriadas. O que acontece depois de um teste que possa dar resultados insatisfatórios, seja por via da negação da hipótese formulada ou mesmo confirmação não infofismável, é simplesmente a sua repetição. A análise quantitativa não oferece uma resposta inicial exata, não descreve nem explica o fenômeno ou processo ou problema que estamos tratando, ao primeiro ensaio. Posta em termos de uma linguagem mais simples e mais antiga, a dialética entre o modelo e realidade não é mais que tentativa e erro, apenas guiada por formulações sensatas e progressivamente corrigidas em testes consecutivos. No caso particular da escolha de lugares, seja a partir de unidades que por qualquer razão sejam não passível de desagregação para fins de análise (e por definição apenas susceptíveis de agregação), seja teoricamente a partir do indivíduo que constitui a unidade mais elementar de tomada de decisão no contexto das ciências sociais, o problema crucial é o nível de resolução do problema (colocada de novo a questão em termos sistêmicos), isto é, como agregar. Em termos de inferência estatística é como inferir em um nível de resolução tal, que estas inferências possam se aplicar ao nível seguinte ou vice-versa. Esta é a essência das duas falácias (ecológica e individualista) e, conforme se comprova na série de artigos do livro de DOGAN (8), os mais acesos debates ainda se situam no terreno das "similaridades e diferenças de modelos sociais e políticos no plano individual, da comunidade e nacional". pp. 80.

Agora exemplificando a partir das diferentes análises realizadas das 209 cidades e dos dois conjuntos regionais. Será que as dimensões do sistema urbano brasileiro são as identificadas no conjunto nacional (mesmo abstraindo-se a questão já proposta de validade apenas a termos regionais e não para o sistema inteiro?) Ou até mesmo questões como as que foram sugeridas inicialmente (as cidades de dez mil habitantes) são relevantes e deveríamos primeiro fazer um teste para se identificar o nível correlação do conjunto de variáveis para cidades de um determinado tamanho na área desenvolvida e diferente na área subdesenvolvida? A dialética é, como afirma BERRY, infundável e os temas inexauríveis. O consolo é que a ciência persegue a verdade por aproximações sucessivas e, assim, cada verdade é absoluta porque acreditamos nela e relativa porque eventualmente mudaremos de opinião, à luz de novas evidências empíricas ou até mesmo sob o ângulo de um novo paradigma da ciência.

O propósito de toda esta discussão não é, absolutamente, obscurecer as inegáveis vantagens do processo de quantificação e matematisação nas Ciências Sociais de um modo geral, e na Geografia em particular, mas apenas o de mostrar os problemas que a sua utilização

levanta. Como já assinalamos e muito claramente confirma ALKER em artigo no livro editado por MATTEI (8) pp. 70, “O raciocínio matemático pode acrescentar apenas formas vazias e tautológicas ao conteúdo de nosso argumento. Apesar disto, a clareza, a certeza, a generalidade abstrata e o potencial para a complexidade desta forma de expressão freqüentemente nos leva, ou ajuda a levar, a surpreendentes e valiosas descobertas”.

2. b) A escolha de variáveis

Expostos os principais problemas associados à escolha de lugares a serem considerados unidades observacionais no processo de análise fatorial, o segundo ponto é de igual importância — o da escolha de variáveis — e é claro que ele não é independente do outro e muitas das implicações da escolha de lugares se refletem na escolha de variáveis e vice-versa.

Entretanto é neste particular que o problema de teorização (aí implícita a idéia de formulação de hipóteses) é fundamental porque, ainda que mais trabalhoso, poderíamos (no caso do Brasil) utilizar em uma análise todas as unidades observacionais que se qualifiquem como cidades, segundo definições formais do tipo sede de município. No caso de variáveis o problema é diferente porque seu número pode ser infinitamente grande e se não relevantes podem obscurecer ainda mais as correlações existentes.

Ainda aqui procuraremos utilizar o exemplo das 209 cidades e 59 variáveis e as hipóteses que precederam a escolha das mesmas. E começaremos por discutir possivelmente a mais óbvia das hipóteses sobre dimensões básicas de um sistema de cidades, que é seu tamanho. Vamos aceitar, sem discussão, a hipótese de que tamanho não é uma medida singular — obtida da forma mais simples que seria o número de habitantes — mas uma medida compósita, constituída de um número de variáveis para refletir efetivamente o tamanho, são apresentadas como um valor absoluto (que tanto pode ser o número absoluto como o percentual que uma determinada cidade representa do total de unidades analisadas). O primeiro problema começa com o fato de que adotamos as cidades de mais de 10 mil habitantes (portanto a partir de um tamanho singular) e vamos, a partir daí, analisar não só seus tamanhos compósitos, como outras dimensões nas quais valores qualitativos dos indivíduos são diferentes (sua capacidade aquisitiva, por exemplo, que incidirá sobre variáveis relativas e indicadores econômicos), fato que já mencionamos como relevante no contexto da seleção de unidades observacionais.

Preliminarmente é preciso conceber o que estamos descrevendo aqui como tamanho compósito: trata-se de um conjunto de tamanhos diferentes um do outro, mas que juntos formam aquele tamanho compósito. Entretanto se colocarmos duas variáveis, que ao invés de representarem duas combinações lineares semelhantes mas não iguais, representando assim uma mesma equação da reta, vamos criar um problema a seguir; em primeiro lugar porque não estaríamos acrescentando nada ao conteúdo do tamanho e sim apenas dando maior importância quantitativa ao mesmo, sem acrescentar dimensão nova semelhante mas não igual. Em segundo lugar uma matriz em que duas colunas sejam redundantes, quer dizer, uma mesma equação da reta satisfaz as duas colunas, é uma matriz singular e as matrizes singu-

lares não têm inverso. Para poder calcular os *scores* das cidades a matriz de *loadings* tem que ser invertida para efeito de cálculo dos coeficientes correspondentes ao mesmo tempo a cada variável e no fator como um todo; se a matriz com variáveis redundantes for utilizada, os *scores* não poderão ser determinados. Este determinante raramente é igual a zero, mas é tornado zero por aproximação e pode variar de programa a programa, mas em geral as variações não são muito grandes e alguns deles permitem um grau mais elevado de redundância. Um exemplo disto é que o programa que vinha sendo utilizado e que não continha o teste de determinante calculava os *scores* de muitas das análises que fizemos aqui, inclusive o desta de 209, cujos exemplos usamos, mas nem a do OSIRIS nem a do SPSS, que estão implantados no sistema computacional da Fundação IBGE, conseguiram calcular os *scores* destas 209 cidades, considerando a matriz singular. A significação desta mecânica do programa é que ela pode orientar a eliminação de variáveis que se repetem, sem acrescentar nenhuma informação adicional e que portanto são inúteis.

Mas deixemos de lado o problema e vamos continuar analisando o conjunto de variáveis que devemos reunir para definir o tamanho funcional, compósito, das cidades. Dois tipos de considerações são relevantes: a primeira diz respeito estritamente ao tipo de variáveis que serão hipotetizadas como gerando o tamanho funcional como o estamos concebendo e outras que podem estar associadas ao tamanho funcional, mas não criando um tamanho. A segunda diz respeito a um balanceamento geral do conjunto de variáveis, com cada conjunto que vise definir uma dimensão hipotetizada, contendo um número de variáveis proporcional à sua importância. Exemplo disso é que podemos considerar que o tamanho é duas vezes mais importante que o nível de desenvolvimento e urbanização e as variáveis de tamanho seriam duas vezes mais numerosas.

Um exemplo do primeiro caso e que reflete as duas componentes do tamanho funcional das cidades no plano nacional e nos dois planos regionais, é que doze variáveis entre as 59 foram consideradas para definir tal tamanho funcional (portanto perto dos 22% que tal fator explica); em relação às 57 cidades o conjunto todo de variáveis produziu 14 fatores contra 11 na análise de 152 cidades e, ainda assim, a importância do primeiro fator na segunda foi ligeiramente inferior ao da primeira, a despeito de ser um entre onze e não um entre quatorze. Os dois percentuais de explicação foram de 26,16% para as 57 cidades e 23,58% para as 152 cidades; é que no primeiro caso, variáveis associadas, como já foi assinalado no início do presente trabalho, mostram que tamanho e desenvolvimento são dimensões independentes apenas para o conjunto de 152 cidades, mas não para as 57 e elas aumentam o poder de explicação do fator. JANSON, em artigo da mesma coleção já mencionada, editada por DOGAN e ROKKAN, mostra que "Uma cuidadosa seleção de variáveis aumenta as chances de obtenção de resultados úteis, de forma considerável. Em pré-testes ou outros estudos exploratórios a seleção é tentativa, mas de outra forma nenhuma variável deve ser incluída exceto por razões teóricas ou técnicas especiais. Se hipóteses definidas não estão sendo testadas, hipóteses implícitas de relevância ou importância para certos campos ou aspectos a serem cobertos podem ser seguidas" pp. 317.

A hipótese implícita na seleção das variáveis de tamanho funcional é a de que elas cobrem os diferentes setores desta dimensão e, portanto, a sua importância relativa é esta. Mas ainda há um problema importante a considerar, no que diz respeito à medida de tamanho

funcional, que é relevante por si mesma, mas provavelmente mais importante no contexto de medidas que implicam em valores absolutos e que é a da utilização de dados brutos ou transformados. Fizemos inicialmente uma análise de algumas variáveis, no que diz respeito aos seus valores médios e ao desvio padrão, ressaltando o significado da escolha de variáveis em diferentes contextos: ora no contexto de uma variável que tenha uma distribuição normal ao longo do conjunto de cidades, portanto com média significativa e desvio padrão pequeno, ora com variáveis que sejam altamente discriminatórias, do tipo das que mencionamos (ensino superior, esgoto ou indústria pesada). Quando se trata de variáveis que visam dar uma dimensão de tamanho absoluto, ainda que normalizadas, os desvios são grandes e mesmo a transformação logarítmica não elimina a distorsão. Por exemplo, consideradas as 209 cidades, e usada a transformação logarítmica, a média para a variável referente a número de pessoas ocupadas na indústria é de 0,31 e o desvio padrão 0,53, referente a número de veículos a média é de 0,35 e o desvio é de 0,55; enquanto isso outra variável referente à população a média é de 0,69 e o desvio é de 0,66, o mesmo ocorrendo com as duas variáveis referentes ao pessoal ocupado em comércio e serviços, com médias e desvios de 0,54 e 0,65 e 0,66 e 0,63, respectivamente. Teríamos aí coeficientes de variação decrescentes a partir de pessoal ocupado em indústrias ou número de veículos, que são discriminatórias para população e pessoal ocupado em serviços que são as duas menos discriminantes e pessoal ocupado no comércio que é intermediária. Ainda aí se constata o potencial destas estatísticas para a construção de explicações do processo. É bastante compreensível que, em um país como o Brasil, em estágios iniciais de desenvolvimento, população e serviços sejam variáveis mais universais que, até mesmo, comércio e, certamente, muito mais que indústria ou automóveis.

Tanto isto é verdade que nas duas análises de 57 e 152 cidades, a universalidade das variáveis população e serviços, e mesmo comércio, permanecem quase que com coeficientes de variação iguais, mas a referentes a automóveis ele é maior no sistema de 57 cidades (na periferia), onde os contrastes de capacidade aquisitiva são mais acentuados, ao passo que a relativa a estabelecimentos industriais e coeficiente de variação é maior nas 152 cidades (no núcleo básico desenvolvido), onde verdadeiras regiões industriais se formam, em forte contraste com núcleos urbanos com funções centrais importantes e que se distribuem com muita regularidade, destacando-se suas funções de comércio e prestação de serviços. Isto é claramente indicado pelo fato da variável relativa à indústria ter coeficiente de variação maior, ao passo que as referentes a indicadores econômicos (água, esgoto, telefone, automóvel por mil habitantes, etc.) têm coeficiente de variação bem menor, o que não ocorre na área subdesenvolvida.

As análises realizadas e comparadas entre si tomaram como base os números absolutos e relativos sujeitos à análise fatorial, ou uma transformação logarítmica de todos eles e esta transformação teve o objetivo de mostrar como se pode obter, inicialmente, uma estrutura de relações mais compacta que a sem transformação (13 fatores e 85,58% de explicação nas 57 cidades e 12 fatores e 77,59% nas 152 cidades na análise sem transformação), com 14 fatores e 87,22% nas 57 cidades e 11 fatores com 77,73% nas 152 cidades, além de ter-se um maior grau de linearidade entre as mesmas, indicada pelos desvios padrão sempre menores. A transformação obedece mais às premissas do método que o uso de variáveis não transformadas, mas isto não significa que a trans-

formação mais adequada seja a logarítmica. É possível que o mais adequado seja o uso de transformação apenas nas variáveis absolutas e que a melhor transformação nem seja a logarítmica, embora esta seja uma das mais usuais.

Por outro lado, ao calcular-se os *scores* das cidades em cada um dos fatores, especialmente no de tamanho funcional, o uso da transformação logarítmica reduz enormemente as diferenças de tamanho entre os grandes núcleos, criando, quem sabe, uma imagem distorcida das diferenças entre os vários tamanhos funcionais. Um exemplo disso é que São Paulo e Rio de Janeiro apresentam tamanhos funcionais, respectivamente, de cerca de 140 e 100 na análise das 209 cidades, sem transformação, comparando com tamanhos poucos inferiores a 30 de Porto Alegre e Belo Horizonte. Usando-se os valores transformados, o tamanho de São Paulo e Rio passa a pouco menos de 90 e pouco menos de 80, enquanto que Porto Alegre, por exemplo, passa a pouco mais de 50. É preciso um esforço de imaginação para conceber as diferenças entre tamanhos como resultantes das diferenças do logaritmo do tamanho (embora se possa usar o artifício de a seguir calcular-se o tamanho real de volta), após operações matemáticas em que a maior linearidade era requerida.

Discutidos os problemas essenciais relativos à escolha de variáveis no contexto de uma dimensão óbvia, como é o tamanho funcional, não cabe, nos limites do presente estudo, analisar as diferentes hipóteses relativas a outras dimensões do sistema urbano, tais como nível de desenvolvimento e urbanização, ou estrutura industrial, comércio e serviços (centros especializados ou localidades centrais), a não ser no mesmo contexto em que foi discutido o fator tamanho: a inclusão de variáveis deve obedecer a um critério específico, seja a título de pré-teste e procura de regularidades que possam dar origem a hipóteses a serem testadas, em que uma variada gama de variáveis é incluída, cobrindo diversos aspectos da vida urbana (caso em que as hipóteses estão implícitas), seja sob o ângulo de uma série de hipóteses definidas de forma clara e precisa. É este o caso de duas dimensões explicitamente procuradas e definidas no conjunto de variáveis, que descrevem uma o nível de desenvolvimento e *status* econômico social, e a outra a especialização funcional (estrutura industrial ou comércio e serviços). Com o propósito de testar explicações prévias de certos fatores, um certo número de variáveis pode ser incluída, como, por exemplo, a distância para São Paulo e Recife, ou o número de centros num raio de 100 e 200 quilômetros, com o propósito de testar a hipótese de que uma rede urbana mais densa está associada a maior proximidade de São Paulo, ao lado de níveis de desenvolvimento maior e de uma população com uma estrutura etária menos caracterizada por um elevado índice de participação de pessoas com idades inferiores a 14 anos e, portanto, fora da força de trabalho. É curioso observar que esta última variável apresenta um coeficiente de variação menor na área subdesenvolvida do que na área desenvolvida e, na estrutura de correlações, ela aparece no fator que define a estrutura comércio/serviços, mas de uma forma diferente. Na análise das 209 cidades o fator 3 identifica a estrutura bipolarizada indústria/comércio/serviços, obtida através de variáveis que refletem relação pessoal ocupado na indústria/pessoal ocupado nos serviços ou no comércio. Assim este fator define centros efetivamente de comércio e serviços dos centros industriais. Já o fator 6, ao identificar os centros de comércio e serviços, o faz através do percentual de pessoal ocupado no comércio e serviços, como uma proporção da população urbana total,

correlacionado ao mesmo tempo com maior distância de Porto Alegre e proporção de pessoas com menos de 14 anos mais elevada. O exemplo mostra bem como a escolha de uma variável derivada, como a indicada no fator 6, pode mostrar que os centros de comércio e serviços, mais na direção do Nordeste, têm uma relação pessoal ocupado/população urbana relativamente baixa, embora fosse de se esperar que assim como as variáveis referentes a pessoal ocupado na indústria em relação à população urbana ficou correlacionada no fator bipolarizado indústria/comércio-serviços, esta também ficasse apenas com sinal diferente. O fato de não ter ficado indica uma defasagem bem maior entre nível de emprego na indústria em relação à população urbana e seu correspondente em comércio e serviços, indício claro de desemprego significativo, embora correlação com população de 0 a 14 anos indique também uma estrutura etária jovem. É claro que estas variáveis ainda são insuficientes para tais inferências, mas indicam o caminho para inclusão de outras mais específicas. É esta a função da inclusão de variáveis deste tipo, que têm um caráter mais explicativo que descritivo, o explicativo aí no sentido de interpretação das correlações encontradas.

Semelhantermente as distâncias para São Paulo, Porto Alegre e Recife foram incluídas com o propósito de testar uma hipótese de que níveis de desenvolvimento decrescentes são encontrados a partir de São Paulo, principalmente, hipótese inteiramente confirmada, uma vez que esta variável apareceu com correlação relativamente elevada no fator 2 da análise, que indica o nível de desenvolvimento e urbanização. Na análise das 152 cidades, portanto da área desenvolvida, a distância para São Paulo deixou de aparecer com correlação significativa, o que mostra bem seu comportamento em dois níveis, relevante como uma definição da função decrescente de desenvolvimento, a partir de São Paulo no plano nacional, mas irrelevante no plano da área desenvolvida. Voltaremos a discutir o papel destas variáveis mais adiante, em outro contexto, significativo do ponto de vista de definir tipologia e regionalização. E assim muitos exemplos ainda poderiam ser citados, neste ou em outros estudos feitos do sistema urbano, que tem o sentido de adicionar poder explicativo ao processo analítico, que é essencialmente descritivo. Aliás a própria média e os desvios padrão para as três variáveis são bem sugestivos de que os dois sistemas são pouco articulados, com um extenso vazio de cidades entre o núcleo básico de cada um dos dois. Na análise das 57 cidades a média na variável distância para São Paulo é de 4,24, enquanto que o desvio padrão é de apenas 0,35, indicando quase uma invariância das distâncias, enquanto que a distância para Recife, de seu próprio sistema urbano tem uma média de 1,70 e um desvio padrão de 0,85; já na análise das 152 cidades a média para São Paulo é de 3,51 e o desvio padrão é de 0,68, enquanto que a distância para Recife tem média 3,11 e desvio padrão de 0,24. Quer dizer que as distâncias internas em cada sistema são menores e sua variação é maior, enquanto que as distâncias entre sistemas é maior, mas a variação é menor, o que de certa forma é um índice de não articulação entre os dois sistemas.

3. O problema da explicação

Mencionamos inicialmente que duas decisões arbitrárias deviam ser tomadas em um estudo que utilizasse análise fatorial como técnica analítica. Entretanto existe uma terceira que, a rigor, não é propria-

mente arbitrária (na realidade as duas outras precisam estar igualmente baseadas em uma concepção teórica), pois está ligada à própria formulação teórica do modelo que queremos testar (seja por uma forma explícita, seja por uma forma implícita no simples fato de escolher um certo número de variáveis). Esta decisão é a de dar nomes aos fatores, nomes que devem estar consistentes com seu conteúdo em termos de uma estrutura de relações e com a explicação que procuramos.

Não deve sofrer a menor dúvida a noção de que análise fatorial é puramente um processo descritivo, em forma quantitativa, de uma estrutura de inter-relações entre um número de variáveis, para determinado número de observações. O seu significado estrito, válido somente para o contexto em que é feito, foi bastante enfatizado no capítulo anterior. Então, como explicar? Diz BERRY (9) “Para entender o como e porque da ecologia fatorial é necessária a perspectiva da filosofia fenomenológica. A essência da perspectiva fenomenológica, não excluindo a fenomenologia transcendental, é a premissa de que conhecimento refletivo pode ser derivado somente dialeticamente, através do *interface* do mundo de nossas experiências nativas, com a atividade estruturadora de nossas diferentes orientações conceituais e perceptuais. A dialética reside na estrita correlação entre o mundo como nós o conhecemos, as premissas práticas e teóricas e os atos que usamos para idealizá-lo. Assim a posição filosófica nos leva a reconhecer que todos os conceitos das ciências sociais estão, em última instância, baseados no mundo de nossa experiência social que, por sua vez, é organizado em modos típicos pela linguagem cotidiana que nós e os autores sociais que estudamos, usamos juntamente com os hábitos perceptuais, que nós e eles aprendemos. Mas o “mundo vivo” somente nos aparece como tal, quando recuamos de nossas crenças e envolvimento, refletimos sobre o *interface* entre as complexidades da experiência concreta e as abstrações seletivas de nossos interesses práticos ou teóricos”. pp. 214 e 215.

Pareceria inconsistente afirmar, a esta altura, que a análise fatorial representaria, neste estágio, o mundo concreto e complexo, através da mensuração de uma variada gama de inter-relações entre variáveis, para um determinado número de lugares. A escolha de umas e de outros (variáveis e lugares) é feita com base nos nossos conceitos de idealizações do mundo real. E é aí que, ainda segundo BERRY, se apresenta o dilema filosófico, “se o mundo real e nossa idealização do mesmo — o objeto e o sujeito — são estritamente correlacionados, como podemos chegar a saber como é este “mundo real”, como ele realmente é? Quando podemos saber quais são os dados brutos (originais), quando os estamos estruturando no próprio ato de conhecê-los? Como podemos superar o fato de que a estrutura idealizadora de uma ciência cria uma moldura que filtra e elimina certos tipos de experiência e que, ao mesmo tempo, focaliza os dados retidos em padrões especificados — fato que é especialmente válido quando sabemos que a orientação básica é tradicionalmente teorização apriorística do tipo positivista?” pp. 215 op. cit.

Em síntese o problema da explicação está fora do contexto da análise fatorial. Ela estrutura os dados, organiza-os em padrões específicos determinados pelas relações entre os mesmos, e estes padrões podem iluminar a explicação que procuramos. Mais facilmente ainda, quando sucessivas análises, transversais e em diferentes níveis de generalização, e longitudinais, em diferentes momentos de tempo, venham

a mostrar seqüências repetitivas de padrões, construídos de variáveis que ao mesmo tempo, definam as dimensões básicas e de outras que estabeleçam relações de natureza estrutural.

Voltemos a exemplificar e usando um que já foi mencionado, relativo à estrutura do fator Tamanho Funcional das cidades, no plano nacional, entre as 57 cidades da área subdesenvolvida e as 152 da área desenvolvida. As variáveis que compõem a estrutura deste fator nas 152 cidades são estritamente indicadoras de tamanho, com exceção apenas de uma parcela da estrutura ocupacional (médicos e outras profissões liberais), o que pode ser constatado nas tabelas 2 e 3 ao mesmo tempo; na tabela 2 aparecem as variáveis que se correlacionam entre si e formam o fator propriamente dito; na tabela 3 podem-se ver os *scores* de cidades grandes e pequenas com igual nível de desenvolvimento no grupo das 152 cidades, o que atesta bem esta independência. Já no que diz respeito às 57 cidades, observa-se na tabela 2 que os componentes do fator I não são somente variáveis de tamanho, mas também outras associadas ao processo de desenvolvimento. A idéia que se forma desta associação é a de que nas áreas subdesenvolvidas as cidades grandes são também as de maior nível de desenvolvimento, o que também se pode ver facilmente na tabela 3.

Vários exemplos que foram utilizados na análise do problema, escolha de variáveis, podem ser repetidos aqui; quando se colocou uma variável referente à percentagem de pessoas com menos de 14 anos, o objetivo era o de testar uma hipótese explicativa de associação de uma estrutura etária com predominância de pessoas naquela faixa, com variáveis indicadoras de diferentes aspectos da estrutura econômica. Para o conjunto do Brasil e quase da mesma forma para as 152 cidades, ela está associada às variáveis relativas ao pessoal ocupado no comércio e serviços, em relação à população urbana, indicando, de alguma forma, que as cidades com estrutura comercial e de serviços (um pouco mais as primeiras que as segundas) contêm uma parcela importante de sua população na faixa etária inferior a 14 anos, mesmo considerando que o desvio padrão em relação à média, para esta variável 0 a 14 anos, é bem baixo. Por outro lado, o fato de que esta correlação surgiu com as duas variáveis comércio e serviços em relação à população urbana e não em relação ao pessoal ocupado em indústria e no comércio ou nos serviços, é uma indicação bastante sugestiva de que estas são as cidades “inchadas” da área subdesenvolvida, pois se tal não fosse, a correlação teria ocorrido nas variáveis indicadoras de dicotomia na especialização funcional.

Por outro lado, nas 57 cidades do Norte-Nordeste, a variável idade 0 a 14 anos aparece correlacionada de forma inversa, com densidade da rede urbana, e menor densidade de população de um lado, e com maior distância para Recife de forma positiva. Quer dizer, as cidades distantes de Recife, onde a rede urbana é menos densa, são aquelas em que a população com menos de 14 anos alcança os percentuais maiores, embora ainda aí o desvio padrão, em relação à média, não seja elevado.

Como se vê algumas variáveis podem ser introduzidas para gerar explicações, embora muitas vezes gerem também necessidades de novas comprovações, como é o caso da inferência relativamente a cidades “inchadas” da área subdesenvolvida.

Outro aspecto a exemplificar, em termos de explicação, ou pelo menos de uma associação estrutural que auxilia a explicação, é a de variáveis como distância para São Paulo e maior densidade da rede

urbana, tendendo a significar que o sistema urbano das proximidades de São Paulo é mais denso, quer dizer, as distâncias entre centros de diferentes tamanhos é menor, associado, em outro fator, ao maior nível de desenvolvimento destas cidades e à sua estrutura industrial. E pode-se levar a explicação mais longe, pois estas mesmas cidades têm *scores* altos, ao mesmo tempo no fator relativo à estrutura industrial e no de desenvolvimento, enquanto que as do oeste paulista, por exemplo, um pouco mais distantes de São Paulo, embora com *scores* altos no fator desenvolvimento, têm *scores* baixos na estrutura industrial e altos na de comércio e serviços.

Em suma, o que pretendemos indicar com este exemplo é que a análise fatorial fornece a explicação, porém somente na medida em que colocamos variáveis que tenham conteúdo explicativo no contexto que estamos conduzindo a análise — quer dizer sob a forma de uma hipótese — e na medida em que ela confirma a hipótese.

Com isso estamos evitando, deliberadamente, a discussão do problema filosófico, seja em termos de o chamado “contexto da descoberta” com sua conotação heurística ou mesmo pragmática, seja em termos de uma teorização “a priori” do tipo positivista. Embora seja difícil separar metodologia de filosofia — uma é forma, outra é conteúdo — e pelo menos adotando a linha analítica seguida por BERRY em seu artigo já mencionado (9) a análise fatorial é um instrumento da dialética entre os dois mecanismos: o *a priori* e o *a posteriori*, entre um sistema rigorosamente dedutivo e outro indutivo. A controvérsia filosófica continua, e quem sabe ela mesma não faz parte da dialética, pois HARVEY (10), em recente resposta a um dos críticos de seu livro *Explanation in Geography*, diz que “o único curso admissível é juntar forma e conteúdo, é usar forma que seja coerente com o conteúdo”, pois não fazê-lo, diz ele, seria tentar fazer a eterna pergunta de quem veio primeiro “se questões de conteúdo são anteriores às de forma”. pp. 325.

Esta colocação nos conduz, agora, à última parte da análise de problemas ligados à utilização de análise fatorial na Geografia, em particular nos estudos urbanos e urbanos-regionais, que é a de, por via de seus resultados, gerar um agrupamento de unidades observacionais: uma tipologia e/ou uma regionalização. É o que veremos a seguir.

4. Análise fatorial, tipologia e regionalização

A análise de um sistema de cidades (ou de forma mais ampla de um conjunto de unidades espaciais) acaba por levar, de uma forma ou de outra, à identificação de similaridades e diferenças, portanto na essência do método geográfico tradicional. Os dois problemas clássicos na identificação de unidades, com um grau especificado de similaridade, são os de regionalização e tipologia. Em suma giram em torno do conceito de região e de tipo. Em termos de metodologia quantitativa o problema se enquadra propriamente, tanto nos métodos como nos conceitos da teoria dos conjuntos. Tanto um tipo como uma região podem ser considerados agrupamentos de unidades que pertenceriam a um conjunto, (claro que o processo inverso de partição, que seria a divisão de unidades maiores em outras menores, também se aplica por igual na noção de conjunto). Até mesmo os conceitos associados de conjuntos formados por elementos que possam pertencer a mais de um conjunto (interseção), aplicam-se estritamente ao problema geográ-

fico, tanto de unidades que possam pertencer a uma mesma região ou tipo, quanto às chamadas regiões de transição (que seriam os conjuntos formados por interseção).

A utilização de análise fatorial para produzir um sistema de classificação é aqui entendida em seu sentido lato, isto é, análise fatorial propriamente dita e seus complementos analíticos de agrupamento (na realidade a própria produção de *scores* a partir dos dados originais e da matriz de coeficientes tirados dos fatores é um complemento). Como os algoritmos de agrupamento partem de premissas de ortogonalidade entre as medidas (pois aplicam o Teorema de Pitágoras), ou se considera a análise fatorial um pré-requisito ou se considera o agrupamento um complemento, o que em si mesmo não é relevante. Na realidade existem questões mais significativas, relevantes do ponto de vista teórico, implícitas na utilização destes métodos.

Vamos identificar algumas delas e discuti-las, embora sumariamente, pois todas contém questões teórico-conceituais da maior importância.

A primeira, implícita e explícita na própria indicação feita anteriormente de que a análise fatorial, a rigor, poderia ser tomada como pré-requisito para o agrupamento, é a de que os processos de agrupamento ao usarem o teorema de Pitágoras, na mensuração de distâncias entre lugares (no caso trata-se de distâncias conceituais), adotam a premissa de ortogonalidade como sinal de independência estatística dos fatores porque a distância final é resultante de uma adição das distâncias interfatoriais. O que isto significa é que se houver não aditividade embutida nos fatores, seja pela simples via de multicolinearidade entre variáveis, seja porque muitas vezes esta colinearidade não é nem meramente espacial, nem essencialmente espacial (por ser talvez diferentes resultados de um mesmo processo causal), estaremos produzindo uma distância por adição, na qual uma mesma quantidade está incluída mais de uma vez.

A mais importante conotação destas técnicas de agrupamento é a sua estreita conexão com hipóteses ou teorias, com a conseqüência de que os agrupamentos de unidades confirmem, ou não, a hipótese, formulada desde a escolha dos lugares e das variáveis.

Um exemplo pode ser mencionado no contexto das análises feitas para as 209 cidades: a hipótese básica que presidiu a escolha das variáveis e das cidades é que o modelo espacial brasileiro se enquadra nos termos propostos por Friedman, BERRY e outros, de um Centro e uma periferia. A classificação resultante deveria constituir, se a hipótese é válida, se os lugares e as variáveis foram escolhidos adequadamente, uma classificação de cidades que refletisse a concepção de um centro e uma periferia. Na realidade, embora a análise confirmasse tal hipótese, ela revelou ainda que no interior do núcleo existem bolsões intermetropolitanos, muito de acordo com diversas formulações teóricas indicadas nos artigos de Brian Berry, bolsões estes que apresentam índices de desenvolvimento inferior aos das áreas adjacentes.

Mas além das premissas implícitas e explícitas na própria metodologia utilizada, existem problemas que são de teoria geográfica propriamente dita e que são inseridas na metodologia usada, através de opções e matrizes especiais. A opção no caso é uma de restrição de contigüidade territorial definida no processo de agrupamento, isto é, somente agrupar uma unidade à outra se ela for definida em uma matriz dada no programa, como contígua a esta outra. Esta opção

cria problemas de natureza conceitual importantes, desde a simples identificação de contigüidade propriamente dita, entre lugares com alguma extensão territorial confrontante, até o problema central relativo ao conceito de região ou tipo.

Este problema central pode ser proposto nos seguintes termos: Uma região é uma extensão territorial, de qualquer tamanho, compacta — portanto sem descontinuidade territorial — definida por um grau de similaridade intrínseca especificada, segundo um certo número de características conceitualmente definidas como relevantes, seja para objetivo específico ou inespecífico? (Neste último caso como definir o relevante e o irrelevante?). Em geral esta tem sido a conceituação clássica de região e é por isso mesmo que numerosos algoritmos de agrupamento contém, entre as duas opções operacionais, uma que é da chamada restrição de contigüidade. Esta restrição, em síntese, significa que o mecanismo iterativo interno do algoritmo procura o par de unidades observacionais mais semelhante — isto é, aquele que tenha uma menor distância global entre si — mas agrupa os dois em uma nova unidade observacional agregada, somente se eles forem contíguos, na forma definida em uma matriz especialmente preparada. Em caso contrário o processo iterativo procura identificar sucessivamente pares de lugares mais semelhantes até encontrar um que seja contíguo, agrupando então os dois lugares. Na realidade o que isto significa é que, com este processo, escolhemos uma entre duas alternativas para agrupar: sacrificamos o grau de similaridade — não agrupando os dois lugares mais semelhantes, a não ser que sejam contíguos — para assegurar a formação de uma unidade que tenha continuidade territorial. Quando procedemos a um agrupamento de objetivos tipológicos, esta restrição de contigüidade deixa de ser imposta, e a consequência é que o agrupamento se processa por maximização de similaridade. O mecanismo iterativo dos processos de agrupamento — embora haja variações entre diferentes algoritmos — é o de maximização de similaridade intragrupos e por definição maximização de diferenciação intergrupos. Em termos estritamente numéricos, sacrificamos precisão e detalhe no agrupamento para obtermos compacidade de resultados, o que constitui uma escolha arbitrária, embora subordinada a objetivos específicos. Exemplificando podemos indicar que Caxias do Sul e Joinville são duas cidades que se assemelham muito, em quase todas as análises feitas, pelos seus níveis de desenvolvimento aproximadamente iguais, por suas especializações industriais semelhantes, sendo Joinville mais semelhante a Caxias do Sul que a Blumenau; no caso de tipologia Caxias e Joinville estariam juntas primeiro, no caso da opção de contigüidade (no caso de se tratar das microrregiões respectivas), até se poder obter uma extensão contínua entre Joinville e Caxias do Sul, teríamos que agrupar áreas imensas altamente heterogêneas.

Em termos estritamente operacionais, a questão pode ser colocada como o foi acima, isto é, uma opção entre contigüidade e não contigüidade, rigorosamente uma opção entre regionalização e tipologia, portanto, duas coisas e dois objetivos essencialmente diferentes. Entretanto a questão é que podemos colocar uma pergunta adicional, de natureza conceitual, que implica numa eventual revisão do problema, como foi proposto inicialmente, isto é, em termos de que tipologia e regionalização são duas coisas e dois conceitos distintos. A pergunta adicional é a de que não será a regionalização um subgrupo da tipologia? Em outras palavras, uma região não será uma unidade tipológica que por coincidência seria ao mesmo tempo contígua?

Sublinhamos a coincidência para discutir o problema em dois níveis: o primeiro, embora não completamente independente do segundo, seria o de que tipologia seria uma noção básica de similaridade entre lugares e a região seria um subconceito, desde que seria rigorosamente uma restrição ao processo de produzir uma tipologia, não no sentido de subordinar o agrupamento a uma prévia constatação de contigüidade, mas sim de subordinar a caracterização da região a uma prévia constatação de tipologia, à qual uma característica adicional de contigüidade seria acrescentada em termos meramente descritivos. O segundo nível de discussão do problema é mais transcendental, pois envolve o próprio conceito de região em termos de um determinado espaço. Consideramos este espaço (aqui no sentido de uma determinada localização segundo os conceitos de teoria de localização) algo sobre o qual uma estrutura socioeconômica se estabelece, sem vinculações quase que de ordem genética de natureza causal ou relacional, e então para definir uma região precisamos apenas definir a tipologia socioeconômica e impor sobre ela uma restrição de continuidade ou contigüidade territorial? Ou a localização geográfica — de que a continuidade ou contigüidade territorial constitui apenas uma expressão — não é uma variável exógena no processo de regionalização, mas sim um conjunto de atributos locais que são parte integrante do próprio processo de regionalização, que seria mal descrito, mal definido e mal interpretado no contexto que tais algoritmos de agrupamento o colocam, quer dizer, uma restrição de contigüidade ao processo de tipologia? Quando especialistas como BRIAN BERRY ou JOHN FRIEDMAN falam *space forming* e *space contingent* não é precisamente isso que estão dizendo? Quer dizer, que o espaço não pode ser dissociado do contexto do processo socioeconômico e que para ser melhor compreendido o mecanismo pelo qual ambos interagem para produzir uma região, variáveis locais devem ser introduzidas na análise do processo e a tipologia resultante deverá constituir regiões, na medida da validade dos pressupostos acima e na medida em que eles revelarem uma integração dos dois processos — o socioeconômico e o espacial?

O problema de conceituação de região tem sido muitas vezes questionado em termos de só se considerar efetivamente uma região, um determinado espaço organizado e articulado, de tal maneira que variáveis locais relevantes estejam associadas a variáveis definidoras do processo socioeconômico, quase que como se pudessem considerá-las a nível de insumos em uma eventual matriz de insumo-produto do tipo ecológico-econômico. Isto não exclui a hipótese de regiões não articuladas, do tipo convencional, pois as primeiras seriam associadas ao conceito de *space forming* enquanto que as outras estariam mais vinculadas à idéia de *space contingent*, tudo num contexto de diferentes estágios de desenvolvimento e portanto de integração espacial.

Toda esta discussão nos coloca de volta e sempre ao problema de teorização e hipotetização, que está sempre no cerne de toda espécie de metodologia, simplesmente porque constatamos, a cada momento, teoria e metodologia associadas, como forma e conteúdo na concepção de HARVEY, antes mencionada (10). De volta, dizemos, à questão crucial da exigência de continuidade ou não para a formalização do conceito de região que é, conforme acentuamos, o problema central na diferenciação entre tipologia e regionalização. JOHNSTON (11) discute amplamente o problema e defende vigorosamente a idéia de que "regionalização com restrição de contigüidade simplifica muito e opera contra a eficiência do teste de hipóteses. Não existe base na teoria geográfica, mesmo na relativa a regiões funcionais para o critério

de adjacência. Se uma região é definida como uma unidade compacta e um grupo é produzido com dois subgrupos separados espacialmente, então temos duas ou mais regiões do mesmo tipo, o que não é um fato incomum” pp. 295. T. Czyz vai mais longe (12) ao afirmar que “a Regionalização é um procedimento em dois níveis: parte de tipos para regiões. Primeiro uma tipologia multidimensional das unidades fundamentais é completada e, subseqüentemente, analisando-se a distribuição dos tipos, regiões são delimitadas. O número de regiões resulta do número de unidades observacionais fundamentais, espacialmente contíguas, agrupadas no contexto de um tipo definido”. pp. 116. Quer dizer, regiões são subunidades de uma tipologia, na hipótese de serem elas espacialmente contíguas, por razões intrínsecas à própria análise e não impostas exogenamente. O argumento fundamental de JOHNSTON é que a restrição de contigüidade não testa a hipótese de que áreas homogêneas contíguas existem, mas apenas a de que áreas contíguas existem e que são, até certo ponto, semelhantes.

Não é o propósito do presente estudo uma análise dos conceitos de região e tipos, que daria margem a uma longa e exaustiva discussão em si mesma, mas apenas ressaltar os problemas que se associam ao uso de análise fatorial e seus complementos analíticos no campo das pesquisas geográficas e no que diz respeito à tipologia e regionalização, dois dos mais significativos problemas na geografia. Problemas que obviamente estão sempre presentes na teoria geográfica, mas que são explicitados, de forma lógica e sem ambigüidades, na metodologia, quantitativa exemplificada.

Entretanto, como em outros casos, procuraremos exemplificar, pelo menos parcialmente, o uso de variáveis de natureza locacional espacial, endogenamente ao processo de análise, em vez de procurar introduzir restrições do tipo contigüidade territorial ao processo de agrupamento. Na própria análise de 209 cidades dois tipos de variáveis foram utilizadas, com o propósito de testar certos tipos de comportamento espacial; tais variáveis foram de um lado o número de centros de uma determinada categoria num raio de 100 a 200 quilômetros e de outro as distâncias de cada cidade para São Paulo, Recife e Porto Alegre. O propósito das duas primeiras foi o de testar as relações estruturais entre nível de desenvolvimento, especialização funcional, etc., com a densidade da rede urbana refletida por maior número de centros a 100 e a 200 quilômetros.

É curioso observar, como já assinalamos antes, que para as 57 cidades do Norte e Nordeste, o coeficiente de variação em torno da média das distâncias para São Paulo é extremamente baixo, enquanto que para Recife, embora baixo, é muito mais alto que para São Paulo (8 e 50% do valor da média); o inverso ocorre para as 152 cidades do Centro Sul, que têm baixo coeficiente de variação para a distância para Recife e bastante mais alto para São Paulo, indicando uma espécie de vácuo entre os dois sistemas, principalmente porque a nível das 209 cidades o coeficiente de variação da distância para São Paulo é maior que para Recife, o que mostra que as cidades do sistema estão a distâncias variáveis de São Paulo, mas sempre a distância menos diferenciada para Recife. Por outro lado a distância para São Paulo está associada à rede urbana mais densa (maior número de centros a 100 e 200 quilômetros) e indiretamente a índices mais elevados de desenvolvimento e especialização industrial, ao passo que Recife está associado a estrutura do comércio e serviços. É claro que numerosas variáveis de natureza locacional podem ser utilizadas para testar hipóteses explicativas que envolvam relações de características físicas

do espaço, com fenômenos de qualquer tipo a elas associadas de forma causal. No plano da aplicação de análise fatorial para estudos climáticos, variáveis como altitude, ou distância para o mar, podem ser usadas para testar hipóteses de variação da temperatura inversamente proporcional à latitude e à altitude, de um lado e amplitude crescente da variação térmica anual com aumento de distância para o mar.

5. Conclusões

Um trabalho deste tipo não contém uma conclusão, no sentido formal. Sumariza problemas levantados e na realidade foi este o objetivo estrito do mesmo, desde que não pretendemos ter resolvido problemas que são sempre motivos de viva controvérsia. Na realidade a aplicação de técnicas analíticas de maior precisão e de maior capacidade operacional avivam estas controvérsias, colocando-as em perspectiva própria.

Pela própria disposição dos itens do sumário, a escolha dos lugares e das variáveis, seja pela via de teste formal de hipóteses ou teorias produzidas "a priori", seja como uma etapa preliminar de pesquisa, constitui o problema central crucial da tarefa. Extremo rigor e cuidado são requisitos essenciais ao bom êxito da análise. Na realidade a escolha de variáveis é tão essencial que pode transformar um método descritivo de uma estrutura de interrelações em poderoso instrumento auxiliar da própria explicação científica.

A utilização dos complementos analíticos de agrupamento para produzir tipologia e regionalização são também apontados tanto em sua forma puramente operacional como quanto aos problemas que levanta: o mais transcendental, naturalmente, é aquele que gira em torno do próprio conceito de tipologia e regionalização, provavelmente o mais controverso, mas provavelmente aquele em que, através de uma metodologia quantitativa do tipo indicado no presente estudo, pode ser visto em uma perspectiva talvez não passível de ser observada pela forma convencional de análise geográfica, do tipo cartográfico. O mapa seria a variável locacional (ou variáveis), e ele é imposto, exogenamente, aos dados que estamos analisando. Mas raramente isto é percebido, pois o estamos usando como um instrumento de visualização de uma distribuição e de similaridades de distribuição.

BIBLIOGRAFIA

- (1) FAISSOL, Speridião — “O sistema urbano brasileiro: uma visão ampliada no contexto do processo de desenvolvimento brasileiro”. *Revista Brasileira de Geografia*, 1972, n.º 3.
- (2) ROBINSON, C. W. — Ecological Correlation and the Behaviour of Individuals na *American Sociological Review*, 15 (1950), pp. 351/357.
- (3) MATTEI e ROKKAN — *Quantitative Ecological Analysis in the Social Sciences*, p. 78. The M.I.T. Press, 1969.
- (4) FAISSOL, Speridião — Além do artigo citado também ver na *Revista Brasileira de Geografia*, Ano 32, n.º 4 — “As grandes cidades brasileiras...” e outro no *Boletim Geográfico*, n.º 223, 1972, “Tipologia urbana e regionalização do desenvolvimento econômico...”.
- (5) DOGAN, Mattei e ROKKAN, Steim — op. cit.
- (6) HARVEY, David — *Explanation in Geography*, Edward Arnold, 1969.
- (7) DEUTSCH, Karl W. — On methodological problems of Quantitative Research em “Quantitive Ecological Analysis in Social Sciences” — M.I.T. Press, 1969.
- (8) DOGAN, Mattei e ROKKAN, S. — op. cit.
- (9) BERRY, J. L. Brian — The logic and limitations of comparative factorial ecology — *Economic Geography*, Vol. 47, n.º 2, Suplemento, junho de 1971.
- (10) HARVEY, David — “On obfuscation in Geography, A Coment on Gale’s heterodoxy”, *Geographical Analysis*, julho de 1972.
- (11) JOHNSTON, R. J. — Grouping and Regionalizing: some methodological and technical observations, em *Economic Geography*, Vol. 46, n.º 2, Suplemento, junho de 1970.
- (12) CZYZ, T. — The aplication of multifactor analysis in Economic Regionalization, *Geographia Polonica*, n.º 15, 1968, pp. 115/134.