

O QUE É, COMO SE FAZ

- Auto-estima*, J. V. Bonet  
*Avaliação escolar*, P. Morales  
*Bioética*, F. Lolas  
*Criatividade & redação*, R. Marchioni, 3ª ed.  
*Cultura brasileira*, A. Vannucchi, 3ª ed.  
*Educação cristã no terceiro milênio*, J. G. Roca  
*Educação na família e na escola (A)*, J. S. i López  
*Ensaio filosófico*, A. P. Martinich  
*Ensino na escola*, M. Saint-Onge, 2ª ed.  
*Ética social*, E. Chiavacci  
*Internet na escola*, A. Sobral, 3ª ed.  
*Jejum*, P. Sciadini, 5ª ed.  
*Lectio divina*, T. Hall  
*Método em sociologia (O)*, J.-C. Combessie  
*Motivação em sala de aula*, J. A. Tapia; E. C. Fita, 5ª ed.  
*Oração*, H. J. M. Nouwen, 2ª ed.  
- *Pecado*, X. Thevenot  
*Pedagogia da gestão mental*, A. de la Garanderie; G. Cattani  
*Pesquisa na escola*, M. Bagno, 15ª ed.  
*Preconceito lingüístico*, M. Bagno, 31ª ed.  
*Que preciso saber sobre clonagem e transgênicos (O)*, Q. M. Arantes  
*Relação professor-aluno*, P. Morales, 3ª ed.  
*Universidade comunitária*, A. Vannucchi

JEAN-CLAUDE COMBESSIE

PARTE II      PARTE III  
4.1      E      5.3  
leitura      leitura  
Complementar      obrigatória

# O método em sociologia

## O que é, como se faz

Tradução

MARIA STELA GONÇALVES



Título original:

*La méthode en sociologie*

© Éditions La Découverte, Paris, 1996, 1999, 2001

9 bis, rue Abel-Hovelacque — 75013 Paris

ISBN: 2-7071-3490-2

PREPARAÇÃO: Carlos Alberto Bárbaro

DIAGRAMAÇÃO: Telma dos Santos Custódio

REVISÃO: Maurício Balthazar Leal

Rita de Cássia Machado Lopes

### Edições Loyola

Rua 1822 n° 347 – Ipiranga

04216-000 São Paulo, SP

Caixa Postal 42.335 – 04218-970 – São Paulo, SP

☎: (0\*\*11) 6914-1922

☎: (0\*\*11) 6163-4275

Home page e vendas: [www.loyola.com.br](http://www.loyola.com.br)

Editorial: [loyola@loyola.com.br](mailto:loyola@loyola.com.br)

Vendas: [vendas@loyola.com.br](mailto:vendas@loyola.com.br)

*Todos os direitos reservados. Nenhuma parte desta obra pode ser reproduzida ou transmitida por qualquer forma e/ou quaisquer meios (eletrônico ou mecânico, incluindo fotocópia e gravação) ou arquivada em qualquer sistema ou banco de dados sem permissão escrita da Editora.*

ISBN: 85-15-02595-7

© EDIÇÕES LOYOLA, São Paulo, Brasil, 2004

## Sumário

Introdução .....	7
Primeira Parte	
DESCOBERTA E COLETA DOS DADOS	
I. A observação .....	23
II. A entrevista semidirigida .....	41
III. O questionário .....	55
IV. Sondagens, amostras .....	75
Segunda Parte	
O TRATAMENTO: ANÁLISE E SÍNTESE	
V. Ordenar, classificar .....	93
VI. Classificações e relações estatísticas entre duas variáveis ..	113
VII. Comparar desigualdades .....	139
VIII. Combinatória .....	155
Conclusão: Sociologizar o método .....	175
Anexo: Elementos de estatísticas e de probabilidades .....	179
Bibliografia .....	187

## VIII. Combinatória

Em vez de comparar diferentes tabelas com a ajuda de operações e de índices calculados sobre cada uma, pode-se desejar associar e ligar mais fortemente suas variáveis a fim de determinar com maior precisão suas ligações respectivas. Vários métodos de tratamento permitem combinar entre si variáveis numerosas ( $n > 2$ ). Pode-se desejar comparar sua contribuição seja a uma noção que as subsume, seja às variações de uma variável dependente, seja à configuração de conjunto das ligações. Os objetivos diferem: mais sincréticos quando se trata de hierarquizar e de somar (escalas de atitude), propriamente analíticos quando a questão é diferenciar as influências (análise multivariada), ao mesmo tempo analíticas e globais se se procura determinar a configuração de conjunto das relações (análise fatorial). Mas todos esses métodos têm em comum o emprego de uma elaboração complexa, “sofisticada” dizemos às vezes, que corre o risco de desviar a atenção dos procedimentos e do que eles implicam para focalizá-la sobre a “eficácia” e a “comodidade” de seus “resultados”.

## ESCALAS E ESCORES

### As escalas de atitude

Acumular as informações e construir um índice global que as resuma é o objetivo das *escalas de atitude*. Elas foram utilizadas com freqüência em psicologia social para reagrupar as respostas dadas a um conjunto de questões consideradas “indicadores” de uma mesma atitude (conservadora, nacionalista, autoritária...).

As escalas são hierárquicas na medida em que, qualquer que seja a forma da questão, as respostas são hierarquizadas de um pólo positivo (a resposta manifesta fortemente a atitude que se quer estudar) a um pólo negativo (ela contradiz a atitude). Confere-se então uma nota a cada resposta, a maior sendo dada à resposta mais próxima do pólo positivo. A fim de homogeneizar os princípios de notação (ao mesmo tempo simplificar a tarefa e limitar os efeitos de estrutura inevitáveis se cada resposta se organiza em função de um número diferente de itens), procede-se na maioria das vezes a uma dicotomização das respostas, seja fechando a questão antes de aplicá-la, seja no momento do tratamento. Se se estima que certas questões são mais importantes do que outras, pode-se também conferir-lhes um coeficiente. Para cada respondente, a soma das notas assim conferidas define um “escore” que o classifica em relação aos outros mais ou menos próximo do pólo positivo da escala de atitude. É um escore sincrético: a escala não permite, por exemplo, que se distingam dois escores iguais obtidos a partir de perfis de resposta diferentes.

### Um índice sociológico de “distância global”

Outros objetos, além das atitudes, podem ser tratados por esse método, notadamente as práticas. Em seu já citado estudo (ver capítulo VI) sobre a “escolha do cônjuge”, Alain Girard [1964] construiu um “índice de distância global entre cônjuges” a partir de indicadores de origem, de trajetória e de posição geográficas e sociais.

Fez-se um esforço para obter um índice único que medisse a distância que separa os cônjuges. Ele foi construído analisando-se doze características dos cônjuges:

- nacionalidade,
- local de nascimento (número de habitantes),
- local de nascimento (situação geográfica),
- nível de escolaridade,
- religião,
- local de residência à época do casamento (número de habitantes),
- local de residência à época do casamento (situação geográfica),
- número de localidades em que morou desde o nascimento até o casamento,
- profissão atual do marido e profissão da esposa antes do casamento,

- nacionalidade do pai da cada cônjuge,
- profissão do pai de cada cônjuge,
- profissão atual do marido e profissão do sogro dele.

Essas doze características não são equivalentes e não têm todas a mesma importância em relação ao fenômeno da distância entre os cônjuges. Não pareceu possível, todavia, determinar *a priori* coeficientes de ponderação e foi conferido o mesmo peso a cada característica.

Segundo a maior ou menor distância que separa cada uma dessas características, foi adotado um barema comum, notando seja em 7 pontos, de 1 a 7, seja em 3 pontos, 1, 4 e 7. A nota 1 representa a maior distância, a nota 7 a similitude.

### Eficácia e alcance do método

A eficácia, as dificuldades e os riscos do método mostram-se de maneira bastante clara nesse exemplo sociológico

que aplica o princípio das escalas hierárquicas para a construção desse “índice de distância global”; por isso seria mais exato, no final das contas, chamá-lo de “índice de proximidade”, pois os escores mais altos correspondem às situações de maior homogeneidade.

O texto destaca a dificuldade encontrada para estimar o que deve ser o “peso” de cada indicador e sua contribuição para o índice global: “nem todos são equivalentes” e no entanto é muito difícil “determinar *a priori* coeficientes de ponderação”. A solução adotada aqui é a mais corrente: não ponderar. Mas podemos, com isso, interrogar-nos sobre a pertinência sociológica desse índice em que um terço dos indicadores de distância são de tipo geográfico (4 em 12): cônjuges socialmente muito distantes mas que têm uma origem e uma trajetória geograficamente análogas acabam sendo classificados como próximos.

Os escores decorrem da vontade de passar de características não-numéricas a valores numéricos: eles propõem classificações resumidas, cifradas precisamente, de leitura e de apresentação simples e socialmente eficazes. Mas há um fosso entre essa exatidão e a arbitrariedade das escolhas que, na prática, a construção da escala implica. Em nome de que decidir que cada indicador deve ter a mesma importância, o mesmo peso? Aparentemente técnica, a decisão implica um posicionamento do qual se vê a ausência de fundamento teórico por menos que, por exemplo, em uma escala de atitude autoritária figurem mais questões sobre a autoridade dos pais do que sobre outras dimensões possíveis da autoridade, sejam elas familiares, profissionais ou militantes. E se se decide conceder uma importância maior a certas questões

do que a outras, ou seja, conferir-lhes uma ponderação superior, em que se deve fundar a escolha do coeficiente? É preciso que ele seja duplo, triplo...? Dicotomizar todas as questões, ou ampliar a escala de notação das respostas a algumas determinadas, também depende de escolhas das quais ninguém pode garantir a neutralização dos efeitos de estrutura nem assegurar-se de um fundamento sociológico.

## A ANÁLISE MULTIVARIADA

Mediante construção, a análise multivariada tem por objetivo distinguir e avaliar a contribuição de várias variáveis para uma mesma variável dependente; uma das maneiras de avaliar a força do elo entre duas variáveis de um quadro consiste então em introduzir uma terceira, até mesmo uma quarta, uma quinta. Daí o nome de análise multivariada. Em sentido estrito, ela designa um método de raciocínio associado à construção de quadros com  $n > 2$  variáveis, método que apresentamos aqui. Mas, em sentido amplo, outros procedimentos que têm o mesmo objetivo, como a análise de segmentação que será exposta a seguir, podem ser colocados sob essa denominação.

### *Um exemplo*

O quadro (p. 120), extraído de Clerc [1964] mostrava que o aproveitamento escolar no ensino fundamental variava em função da categoria socioprofissional do pai. Continuan-

do essa análise, o autor constata que a proporção de “bons alunos” aumenta com os rendimentos e o diploma do pai.

Rendimento mensal	% de bons alunos	Diploma do pai	% de bons alunos
400-1000	39%	Nenhum	38%
1001-1400	43%	Escola primária	39%
1401-2000	53%	Diplomas técnicos	47%
2001 e mais	59%	Certificados técnicos	63%
		Ensino secundário ou mais	68%

“Como esses dois fatores estão estreitamente ligados, análises mais finas se impõem para separar o papel respectivo de cada um” [ibidem].

Clerc cruza, portanto, as três variáveis, obtendo um quadro chamado de *três entradas* (podem-se construir quadros com quatro, cinco, n entradas) e que pode ser apresentado da seguinte maneira:

APROVEITAMENTO DA CRIANÇA (% DE “BONS ALUNOS”)  
SEGUNDO O RENDIMENTO DO TRABALHO E O DIPLOMA DO PAI

Diploma do pai	Rendimento mensal				Conjunto
	400-1000	1001-1400	1401-2000	2001 e mais	
Nenhum	42	36	42	—	38
Ensino fundamental	40	37	44	43	39
Técnico	28	44	52	42	47
Ensino médio e além	—	65	68	65	68
Conjunto	39	43	53	59	48

Nos casos do quadro, cada número indica a porcentagem de bons alunos (quando o pai não tem nenhum diploma e o rendimento mensal da família está compreendido entre 400 e 1000 francos, 42% das crianças são boas alunas...).

### Bloquear uma variável

Quando se lê horizontalmente, em linha, vê-se como, *para um diploma igual*, o aproveitamento escolar varia em função do rendimento: a maioria das variações são irregulares e em três casos em cada quatro chegamos a ver — paradoxo! — a taxa de bons alunos diminuir quando se entra na faixa dos rendimentos mais altos. Quando se lê verticalmente, na coluna, vê-se como, *para um rendimento igual*, o aproveitamento varia em função do diploma: a maioria das variações são regulares e (salvo para os diplomas técnicos na primeira faixa dos rendimentos e, em menor medida, na última) “para um rendimento igual, a porcentagem de bons alunos aumenta quando se aumenta o diploma do pai. [...] A ação do meio familiar sobre o aproveitamento escolar parece portanto quase exclusivamente cultural” [ibidem].

### Alcance e limites

Assim, a análise multivariada propõe um caminho lógico na busca da imputação causal. A força principal desse procedimento deve-se ao fato de que ele permite distinguir a contribuição respectiva de duas variáveis para uma única terceira e assim testar as hipóteses de tipo causal sugeridas

pelas tabelas de contingência que cruzam duas variáveis. Comparando o efeito das variáveis, a análise multivariada conduz também a esclarecer os conceitos (ultrapassar a noção de intercambialidade dos índices, por exemplo).

Ela supõe efetivos importantes, uma vez que cada introdução de uma variável suplementar opera uma partição que restringe o efetivo de caso e diminui a confiabilidade estatística. Por outro lado, além de três variáveis, construir e ler os quadros pode revelar-se tarefa difícil.

A análise multivariada requer além disso uma dupla vigilância metodológica. Inicialmente, como em toda comparação entre tabelas de contingência, uma vigilância quanto aos efeitos de estrutura: mediante construção, as cinco modalidades da variável diploma favorecem variações maiores do que as quatro modalidades da variável rendimento; Clerc não proporcionou a si mesmo os meios de controlar esse efeito de estrutura. Em seguida, a análise multivariada corre o risco de "fazer esquecer" as possíveis interações entre as variáveis ativas: Clerc não se interessou pelo fato de que, cronologicamente anterior, o diploma exercia uma influência certa sobre o rendimento (por um lado, o efeito das diferenças de rendimento é um efeito indireto — ou induzido — das diferenças de diploma). Essa questão será retomada a propósito da análise de segmentação.

### A segmentação

O método dito de segmentação pode ser considerado um modo particular de análise multivariada. Ele também tem por objetivo comparar os elos que existem entre várias variáveis ativas e uma mesma variável dependente. Mas ele

se fundamenta em uma *medida* da amplitude da variação que cada variável ativa induz sobre a variável dependente a fim de poder desembocar em uma tipologia.

### Um exemplo

No exemplo apresentado a seguir, Jean-Paul Grémy [1977] utilizou uma medida de desvio; essa prática é freqüente, mas outras medidas poderiam conduzir a outros resultados.

Para controlar os efeitos de estrutura, cada variável ativa foi dicotomizada (reduzida a duas modalidades), mas a concentração da população em uma ou outra das modalidades das diversas variáveis permanece muito desigual.

O objetivo é comparar a influência de várias variáveis sobre a taxa de aprovação no exame final do primeiro ciclo de

TAXA DE APROVAÇÃO NO PRIMEIRO CICLO DA UNIVERSIDADE SEGUNDO:

		Taxa	Desvio
a CSP do pai	popular	42,7	8,4
	média e alta	48,9	
o exercício de uma atividade assalariada	sim	25,9	37,5
	não	58,4	
a idade da primeira inscrição numa faculdade	< 20 anos	61,3	29,8
	≥ 20 anos	31,5	
a obtenção de uma menção no bac	sim	62,2	20,1
	não	42,1	
a seção do secundário	clássico	63	23,1
	moderno	39,9	

literatura na universidade (o CELG, Certificado de Estudos Literários Gerais): a categoria socioprofissional do pai, a seção cursada durante o ensino secundário (ou seja, “clássico”, seção literária com línguas antigas que, no momento da pesquisa, era a seção mais seletiva, ou seja, “moderna”), o fato de ter ou não obtido uma menção no *baccalauréat*, o fato de exercer ou não uma atividade assalariada durante os estudos.

### Um caminho lógico, um gráfico

Em leitura direta, vê-se que é “ter ou não um trabalho assalariado” que induz o maior *desvio* nas taxas de aprovação.

Divide-se então a população estudantil em dois grupos: os estudantes que têm um trabalho assalariado, os estudantes sem trabalho assalariado. Para cada grupo, calcula-se em seguida a taxa de aprovação segundo as variáveis ativas restantes.

Taxa de aprovação dos: → segundo: ↓		Estudantes assalariados		Estudantes não-assalariados	
		Taxa	Desvio	Taxa	Desvio
CSP do pai	popular	26,5	1,7	53,7	1,8
	média, alta	24,8		59,1	
Idade da primeira inscrição	< 20 anos	35,5	11,2	64,1	20,5
	≥ 20 anos	24,3		43,6	
Menção no <i>bac</i>	sim	35,4	12,2	75,8	23,0
	não	23,2		52,8	
Seção do secundário	clássico	28,2	2,7	71,1	20,4
	moderno	25,5		50,7	

Nesta segunda etapa, “ter ou não uma menção no *baccalauréat*” induz o desvio maior entre as taxas de aprovação.

Determinamos e hierarquizamos, portanto, as duas maiores variáveis ativas. Constituem-se, então, quatro subgrupos estudantis:

- com trabalho assalariado sem menção no *bac*;
- com trabalho assalariado e menção no *bac*;
- sem trabalho assalariado nem menção no *bac*;
- sem trabalho assalariado e com menção no *bac*.

Para cada subgrupo, calculam-se as taxas de aprovação segundo as variáveis restantes. E assim por diante.

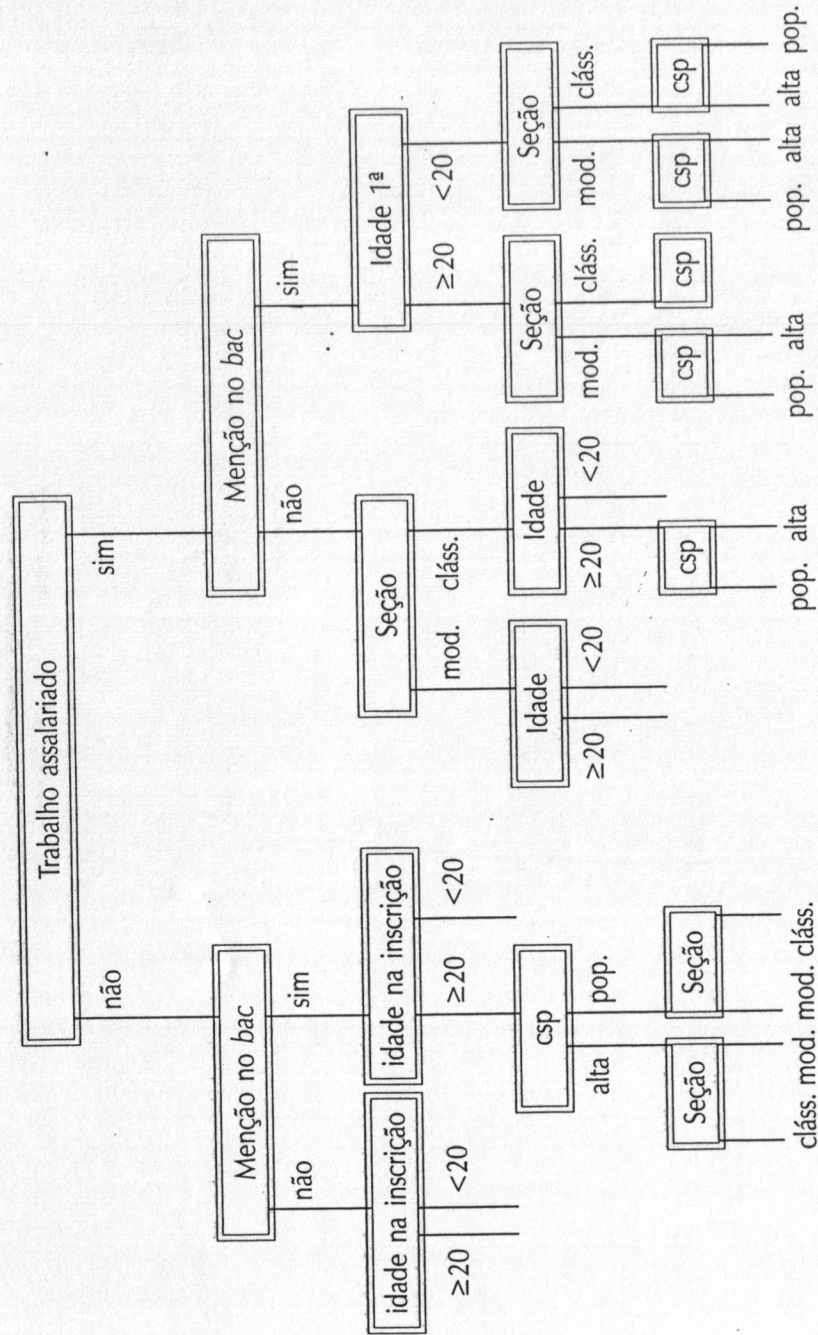
O gráfico da página seguinte esquematiza o resultado definitivo da segmentação. Ele tem a forma de uma arborescência. Lê-se no alto a variável mais ativa, na segunda linha a segunda variável etc. Quando o desvio induzido por uma variável sobre a taxa de aprovação é inferior a três pontos, o trabalho de segmentação não foi continuado.

### Alcance e limites

O exemplo e o gráfico ilustram os efeitos desse método sobre as classificações que ele opera:

- sensação de evidência ligada à representação gráfica;
- dificuldade de controlar o conjunto dos efeitos de estrutura (o efeito do número de modalidades das variáveis foi controlado por dicotomização, mas a concentração de população em cada modalidade permanece desigual);
- neutralização da dimensão cronológica: o protocolo não trata da interação das variáveis ativas entre si e “ignora”





suas *seqüências temporais*. A origem social, por exemplo, é anterior a qualquer outra variável; ora, a análise anterior procede como se a origem social não tivesse nenhuma influência sobre o fato de seguir esta ou aquela seção no ensino secundário, de ter ou não uma menção no *bac*, de entrar mais cedo ou mais tarde para a universidade, de exercer ou não uma atividade assalariada; esta é a razão pela qual se conclui por uma influência muito fraca da origem social sobre o aproveitamento universitário. O método que se seguiu “des-historizou” a análise.

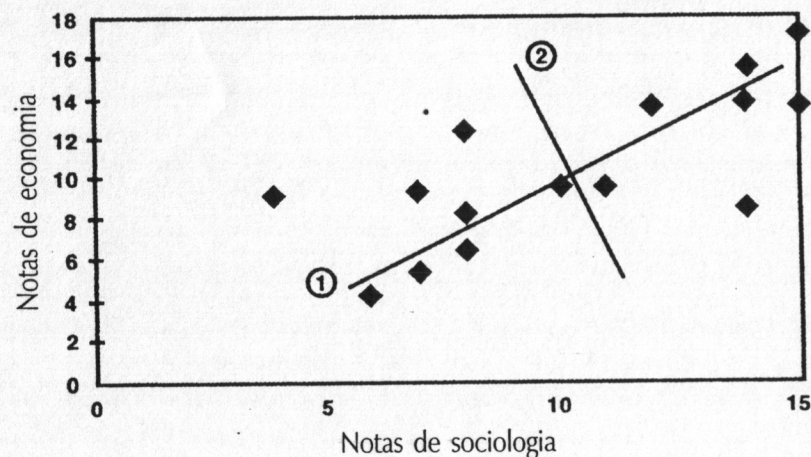
### A ANÁLISE FATORIAL

#### Análise e síntese

A análise fatorial é um dos procedimentos da “análise dos dados” que, assim como a análise de segmentação, coloca em prática “índices” de ligação estatística entre as variáveis. Ela avalia as ligações entre todas as variáveis para hierarquizá-las e/ou propor uma representação da configuração de conjunto dessas ligações. Quer se trate de métodos de análise fatorial, que serão apresentados a seguir, quer de classificação automática, a análise dos dados se destaca por seu duplo objetivo, analítico e sintético, e sua capacidade de dar a “ler rapidamente grandes quadros estatísticos” [Volle, 1980], ou seja, no caso de pesquisas, um conjunto de respostas para muitas questões.

## Nuvens de pontos e eixos fatoriais

Para compreender o princípio da análise fatorial, podemos nos reportar ao exemplo precedente (p. 135) de covariação entre duas séries de notas. A seguir, a “nuvem de pontos” lembra a “correspondência” entre as notas de sociologia e de economia da cada estudante, mas desta vez indicamos o centro delas (a média das notas obtidas em cada prova) e representaremos, passando por esse centro, dois eixos de coordenadas retangulares escolhidas de tal maneira que o primeiro siga a direção mais longa da nuvem. Vê-se que as extremidades desse eixo opõem os “bons em tudo” aos “ruins em tudo” ao passo que o segundo opõe os “melhores em sociologia do que em economia” aos “melhores em economia do que em sociologia”. Interpretar o primeiro eixo é o mesmo que enunciar o fator de covariação positiva que dá sua força principal à nuvem, o segundo indicando as covariações negativas.



## Um exemplo

Tomemos um exemplo mais complexo: como representar as relações entre as modalidades de 24 variáveis “de trajetória social e escolar e de posições na área” dos economistas franceses em 1990? Por meio de um programa de informática adaptado, pode-se calcular tais ligações, mas não imaginar a forma delas no espaço com  $N$  dimensões que a elas corresponde. Por outro lado, pode-se ver bem como, em um espaço com três dimensões, os resultados desses cálculos tomam a forma de uma “nuvem de pontos” cuja largura e cujo comprimento maiores indicam os eixos de variação mais notáveis. Projetada segundo esses eixos sobre o espaço com duas dimensões de uma folha de papel, a “nuvem de pontos” é portanto um resumo das principais variações e “correspondências” cujos dois eixos, chamados “eixos de inércia” ou “eixos fatoriais”, facilitam a interpretação. Assim, na nuvem apresentada a seguir, o primeiro eixo (esquerda-direita) “mostra que os economistas se opõem profundamente em relação às suas propriedades sociais mais gerais” (mais elevadas à direita, mais fracas à esquerda), ao passo que a segunda oposição (norte-sul) “se estabelece entre um pólo de poder intelectual e técnico e um pólo de poder econômico e político”.

## Alcance e limites

*Apelo à vigilância* — Na medida em que a análise fatorial é fundamentada em um uso generalizado dos índices de



em que ela dá aos resultados uma forma gráfica de leitura mais “evidente”.

- *Para um uso correto* — A que práticas esses apelos à vigilância podem levar, no caso de um método que propõe, ao final de procedimentos informatizados difíceis e complexos, uma facilidade de leitura sintética, sinótica, às vezes insubstituível para dar uma visão de conjunto de uma pesquisa por questionários que comportam inúmeras questões?

Pode-se inicialmente sugerir uma comparação sistemática dos efeitos da análise quando se faz variar os diversos procedimentos e parâmetros. Mas esse controle, difícil de ser colocado em prática, nem sempre é realizável.

A regra mais geral para um uso correto da análise dos dados passa por uma reflexão sobre seus usos possíveis *em relação* com os outros métodos da pesquisa (pluralidade dos métodos) e em uma dupla perspectiva de exploração e de verificação.

Figuração global e estruturada, ela pode favorecer a emergência de questões, de “tipos ideais” e de hipóteses cuja validação deverá ser buscada nos outros métodos da pesquisa. Ela é muito particularmente adaptada à elaboração de hipóteses de tipo estrutural e formuladas em termos de campo: os eixos polarizam as hipóteses de leitura e os quadrantes estruturam espaços aos quais essas hipóteses podem conferir uma dimensão imediatamente teorizável.

Após a análise de algumas tabelas de contingência, a análise dos dados pode, como qualquer método de análise, servir para testar hipóteses. Pode-se então limitar (e portanto controlar melhor) as variáveis que lhe são submetidas reten-

do apenas aquelas cujas relações foram previamente estudadas. Se ela faz aparecerem resultados convergentes, as hipóteses são confirmadas; se os resultados diferem, outras questões se apresentam...

A demonstração é tanto mais forte na medida em que ela resulta de operações metodologicamente diferentes. No caso de tratamento estatístico, a análise dos dados será portanto confrontada aos métodos de análise mais simples e logo mais transparentes, por exemplo algumas tabelas de contingência com duas variáveis, permitindo além disso raciocinar sobre os efeitos de estrutura e perceber os espaços críticos induzidos pelas diversas medidas. No caso das informações oriundas dos métodos de coleta com maior potencialidade exploratória, a confrontação entre a análise dos dados e a análise dos relatórios da observação pode ser considerada uma pedra de toque da “verificação” das hipóteses.