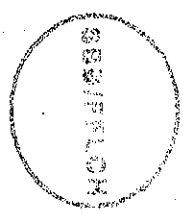
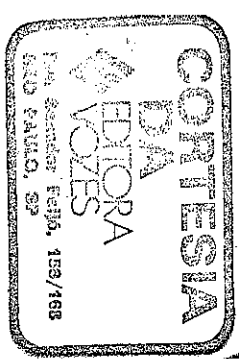


COLEÇÃO  
MÉTODOS QUANTITATIVOS  
EM CIÊNCIAS SOCIAIS

ORIENTAÇÃO DE  
ISAAC KERSTENETZKY

# MÉTODOS QUANTITATIVOS EM SOCIOLOGIA

RAYMOND BOUDON  
Professor da Sorbonne



TOMBO : 114734



BIBLIOTECA DE FILOSOFIA  
E CIÊNCIAS SOCIAIS



EDITORA VOZES LIMITADA  
1971

tão satisfatória quanto possível. É, por exemplo, o que faz Tocqueville quando deduz da noção de centralização administrativa um grande número de traços específicos da sociedade francesa. Acontece o mesmo quando se estuda a mudança: salvo a sociedade relativamente simples, não há quase sentido em querer explicar a mudança pela totalidade da realidade social. A idéia de totalidade é, então, uma noção-limite desprovida de significação operatória.

Mas existe, também, um grande número de pesquisas, que não podem, por isso, ser chamadas "irrisórias" e onde a idéia de *totalidade* está, no entanto, inteiramente ausente: posso perguntar-me se as atitudes acêrca da Escola variam segundo as classes sociais e contribuem para explicar a desigualdade das oportunidades escolares. Posso também me interrogar a respeito das diferenças entre as taxas de suicídio que se observa de uma sociedade a outra sem ter que acusar a sociedade em todos os seus aspectos. Certamente, quando Durkheim faz depender a proporção de suicídios do grau de anomia, êle define uma variável que pode caracterizar a sociedade em seu conjunto. Mas não se pode falar aqui de totalidade.

Em suma, existe um conjunto de pesquisas nas quais a idéia de totalidade pode ter uma significação relativamente precisa, seja porque o objeto possa ser considerado como *exaustivamente inventariável*, seja por ser concebido como *sistema*. Mas há outras em que a idéia de totalidade não representa, manifestamente, nenhum papel e não tem qualquer utilidade.

A idéia de totalidade, portanto, não é mais indispensável em sociologia do que em física. Inútil quando se estudam as leis de choque, ela toma um sentido — através da idéia de sistema — em cibernética. Da mesma forma, ela tem um sentido em Montaigne ou em Murdock, mas quase não o tem em *Le Suicide*, de Durkheim.

## CAPÍTULO II

### OS MÉTODOS DAS ENQUETES QUANTITATIVAS

PODEM-SE definir as enquetes quantitativas como as que permitem recolher num conjunto de elementos informações comparáveis de um elemento a outro. E' esta comparabilidade das informações que permite, em seguida, os recenseamentos e, mais geralmente, a análise quantitativa dos dados.

A condição necessária para a aplicação dos métodos quantitativos é, pois, que a observação repousa sobre um conjunto de elementos, de uma certa forma comparáveis. Na maioria das vezes elementos são indivíduos, mas podem ser, também, grupos, instituições, sociedades ou outros tipos de unidade.

O ponto de partida de uma enquete qualquer — quantitativa ou qualitativa — é, em geral, uma questão do tipo *por quê?* — Por que o suicídio varia segundo época e lugar? Por que se decide votar em tal candidato? Quais são os fatores do absentismo profissional? do anti-semitismo? Por que o clima de uma empresa é melhor ou pior?

Estudaremos neste capítulo a seqüência dos procedimentos que permitem passar de questões deste tipo a suas respostas. No caso em que é possível e se decide utilizar um modo de observação quantitativo, êsses procedimentos aparecem como relativamente idênticos — com variantes próximas — de

uma pesquisa a outra. Distinguiremos os quatro *momentos* seguintes:

- 1) a formulação das hipóteses;
- 2) a construção do plano de observação;
- 3) a construção das variáveis;
- 4) a análise das relações entre variáveis.

### I. A formulação das hipóteses

A formulação das hipóteses pode ser feita *a priori*. Assim, posso perguntar-me se o fato de ter pertencido a um grupo minoritário na infância acarreta a aparição de sintomas depressivos,<sup>1</sup> ou se o fato de ter sido criado numa família autoritária compromete o desenvolvimento da personalidade, ou se as crises econômicas aumentam as taxas dos suicídios, ou se uma organização de tipo *burocrático* acarreta a insatisfação e a ineficácia.

Neste caso, a hipótese é claramente enunciada e se poderá passar diretamente à etapa seguinte.

Em outros casos, é mais difícil enunciar diretamente as hipóteses. Assim, posso ficar impressionado pelo fato de que os suicídios estão muito desigualmente divididos segundo as sociedades, sem estar por causa disso em condições de enunciar hipóteses precisas sobre a natureza do fenômeno. Ou então, se interrogo uma população de alunos para saber se eles já têm intenções precisas a respeito da profissão que desejam seguir, posso ficar impressionado pelo fato de que a proporção dos indecisos varia muito de um estabelecimento a outro ou de uma seção a outra, sem ser capaz de emitir hipóteses precisas sobre as razões dessas variações. Ou então, posso simplesmente me perguntar por que uns já têm intenções firmes e outros não.

<sup>1</sup> ROSENBERG, Morris, Contexte religieux dissonant et perturbation émotionnelle, in Boudon e Lazarsfeld, *L'analyse empirique de la causalité* (cf. Bibliografia).

Duas situações são então possíveis. A primeira é aquela diante da qual se encontrou Durkheim quando empreendeu seus estudos sobre *Le Suicide*. Em razão da natureza do problema ou de limitações financeiras, o sociólogo pode encontrar-se na impossibilidade de recolher por sua própria conta a informação que deseja e é obrigado a se apoiar em dados da *contabilidade social* recolhidos pelos organismos de estatística.

No caso em que lhe é possível recolher uma informação talhada para servi-lo, se a literatura existente não lhe fornece hipóteses suficientemente seguras ou se teme não poder formular corretamente de improviso seu instrumento de observação, procederá ao que se chama geralmente *pré-enquete*. Isto é, procederá a um reconhecimento do terreno, tentando desembaraçar-se de suas idéias preconcebidas ou, como dizia Bacon, de suas prenoções, de maneira a fazer aparecer os fatores ou variáveis explicativas que pesquisa.

Ilustremos isto, por exemplo, pelo célebre estudo sobre *The Authoritarian Personality*. Após a segunda guerra mundial, uma equipe animada pelo filósofo e sociólogo alemão Theodor Adorno, então morando nos Estados Unidos, empreende a análise do anti-semitismo sob um ângulo psicossociológico. No ponto de partida, os autores de *The Authoritarian Personality* dispunham de uma hipótese relativamente geral e imprecisa, segundo a qual o anti-semitismo seria um dos sintomas do que se pode chamar uma síndrome de personalidade. Em outros termos, a idéia era de que o anti-semitismo devia corresponder a um tipo de personalidade determinada de que se tratava de inventariar as características.

Partindo desta idéia, os autores procederam a uma série de entrevistas centradas<sup>2</sup> em um certo número

<sup>2</sup> Sobre os tipos de entrevistas e de questionários, consultar, por exemplo, FESTINGER e KATZ, *Les méthodes des sciences sociales*, Paris, Presses Universitaires de France.

de estudantes, provocando suas reações relativas a temas extremamente variados. Eis curtos extratos do protocolo de um estudante anti-semita:

Vocês são pagos, na Administração, em função do número de pessoas que têm sob suas ordens. E' porque *eles* recriam o maior número possível de pessoas. Só pensam nêles. Não sou materialista (mercenary) bastante para compreender essa attitude. Simplificaria tudo através de uma administração competente... E' preciso eliminar a confusão. O que dirige deve escolher seus colaboradores com cuidado. Atualmente não há autoridade bem definida.

Vê-se destacar do texto um certo número de temas. Tema da recriminação: o sistema administrativo é acusado de confusão generalizada. Tema do egoísmo *dos outros*, acusados de obedecer a razões materialistas e egoístas, de desprezar o interesse geral. Tema da autoridade: a confusão resulta da ausência de autoridade, daí a necessidade de mudar violentamente o sistema.

Estes temas fazem parte da síndrome procurada? Não se pode, naturalmente, decidi-lo, a partir dessa única entrevista. E' possível, na verdade, que estejam ligados ao sintoma "anti-semitismo" nessa pessoa particular mas que a ligação não seja geral. E' possível, também, que a entrevista descreva uma situação objetiva.

Entretanto, uma entrevista como esta ajuda a precisar as hipóteses iniciais. Fornece um certo número de temas de que se verificou em seguida, em uma grande população, que representavam efetivamente os elementos constitutivos de uma síndrome.

Eis um outro exemplo de entrevistas que não foram, na verdade, obtidas em uma pré-entrevista, mas que podem servir para mostrar a utilidade dêsse tipo de informação para a formulação das hipóteses.

Suponhamos que nos coloquemos o problema de saber por que, com inteligência igual, com sucesso escolar igual, certas crianças oriundas de classes desfavorecidas atingem o ensino superior e outras não. Eis aqui dois fragmentos de entrevistas, registrados por J. Kahl.<sup>3</sup> Reproduzem as expressões dos pais de duas crianças de meio humilde que tiveram brilhante sucesso em seus estudos secundários. O primeiro não quer ir mais longe. O segundo está decidido a entrar na universidade.

*Primeiro caso* (o filho quer parar de estudar):

"Eu mesmo nunca fui brilhante. Tudo que desejo é poder ganhar bastante dinheiro para poder viver, sem me inquietar com o futuro... Gostaria que meu filho fizesse melhor que eu... Não o forço... Não sei o que êle preferiria fazer. Além disso, não é problema meu, não tenho que dizer o que eu queria que êle fizesse. Talvez êle não fôsse bastante qualificado. Tentei dizer-lhe que deveria ser médico ou advogado ou qualquer coisa dêsse gênero. Eu lhe disse que deveria estudar inglês e aprender a tratar as pessoas. Êle poderia ser representante... Foi a única sugestão que lhe fiz... Há crianças que têm uma idéia precisa na cabeça e a seguem, mas a maioria daquêles com quem falei pegam o que aparece... Não acredito que um diploma seja tão importante. Quero dizer com isso que, se o senhor se apresenta em algum lugar e o senhor diz "tenho um diploma", isso pode facilitar as coisas mas é tudo em que pode servi-lo".

Veamos, em oposição, a segunda entrevista:

*Segundo caso* (o filho quer entrar na universidade):

"As pessoas que estiveram na universidade parecem mais bem sucedidas. São mais capazes de ter empregos de tipos diferentes. Talvez não saibam mais que os outros, mas sabem como aprender. De um certo ponto de vista, aprenderam a aprender mais facilmente... se têm um emprego que não leva a nada, são capazes de sair dêle ou de mudar...".

<sup>3</sup> "Common man boys, in A. H. Halsey, JEAN FLOUND e ARNOLD ANDERSON, *Education, economy and society*, Nova York, The Free Press, 1961.

O contraste é impressionante. No primeiro caso a importância da instrução é minimizada, o diploma é concebido apenas como cartão de visita. No segundo caso, os estudos são percebidos como fonte de potencialidade. Neste, a situação social é concebida como sendo, na maioria dos casos, resultado do azar. No outro, é representada como derivando de uma conduta de pesquisa racional, tornada mais fácil pela cultura adquirida na universidade. Além disso, a concepção do sucesso é diferente nos dois casos. No primeiro, é concebido como aquisição do conforto material e da segurança. No segundo caso, trata-se de encontrar uma profissão que "leve a alguma coisa".

Em uma palavra, vê-se sair do contraste entre estas duas entrevistas uma hipótese; a de que um dos fatores que contribuem para frear a mobilidade social é a representação do sucesso e das vias de acesso ao sucesso. Para mostrar a validade desta hipótese, seria necessário, então, verificar que o "sistema de valores" enunciado pela primeira entrevista é efetivamente mais típico dos meios humildes.

## II. A construção do plano de observação

Uma vez formuladas as hipóteses, seja *a priori*, seja a partir de uma pré-enquete, trata-se agora de verificá-las. Para isso é preciso inicialmente escolher um plano de observação.

1. Tipos de dados. — Segundo os problemas que se colocam, as possibilidades que se dispõem e diversas outras circunstâncias, o sociólogo é levado a utilizar diversos tipos de dados.

Durkheim, em seu estudo sobre o suicídio, utiliza os dados estatísticos da contabilidade social. Thomas e Znaniecki, em seu monumental estudo sobre o camponês polonês emigrado para os Estados Unidos,

utilizam coleções de correspondência e de documentos biográficos.<sup>1</sup> Analisam fatos do quotidiano em que estão implicados emigrados poloneses, consultam os arquivos dos tribunais e dos comissariados de polícia e se esforçam, a partir dessa documentação múltipla, em destacar os efeitos da transplantação. Em um estudo sobre o funcionamento dos comitês-de-empresá, M. Montclair utiliza atas de reunião.<sup>2</sup> Em uma pesquisa em curso sobre mudanças na concepção que os docentes têm tido do ensino de cem anos para cá, V. Isambert teve a idéia engenhosa de analisar os temas dos discursos de distribuição de prêmios.

Segundo o tipo dos dados utilizados, problemas metodológicos inéditos se apresentam. Assim, a análise dos documentos exige que sejam convenientemente separados os elementos pertinentes do material; encontram-se, então, problemas de *análise de conteúdo*. Os dados estatísticos oriundos da contabilidade social colocam naturalmente o problema da crítica das fontes. Assim, Halbwachs observa que é preciso interpretar com prudência o fato de que a taxa de suicídio da mulher aparece ao nível das estatísticas como mais baixas que a do homem.<sup>3</sup> Na verdade, as formas de suicídio se repartem diferentemente, segundo os sexos: a mulher recorre ao suicídio por afogamento mais freqüentemente que os homens. Ora, é mais fácil dissimular um suicídio por afogamento como um acidente do que um suicídio por arma de fogo.

Qualquer que seja o tipo de material utilizado, o ideal é sempre obter dados que possam ser comparados entre si. Assim, uma das vantagens de recorrer aos discursos de distribuição de prêmios para compreender as mudanças das ideologias ligadas ao en-

<sup>1</sup> *The polish peasant in Europe and America* (1918-1920).

<sup>2</sup> *La dynamique des Comités d'entreprise*, Paris, C.N.R.S., 1963.

<sup>3</sup> *Les causes du suicide*, Paris, Alcan, 1930.

sino é que estes discursos aparecem num quadro institucional determinado, com uma periodicidade regular.

Os dados que evocamos até aqui têm o caráter comum de serem impostos ao pesquisador. Não são construídos por sua iniciativa. Naturalmente não pode ser de outra forma quando se quer analisar uma evolução histórica e em certas outras circunstâncias. Mas o fato de dispor de dados que não foram construídos por ele mesmo apresenta um inconveniente manifesto: não se encontram nêles, necessariamente, as informações que aí procuramos.

É por este motivo que a enquete por questionário ou por entrevista assumiu uma extensão considerável desde a segunda guerra mundial.

Detenhamo-nos, um instante, em um exemplo tomado à sociologia eleitoral para medir as vantagens lógicas da enquete por sondagem em relação aos dados fornecidos pela contabilidade social.

Para explicar a repartição dos votos numa consulta, o sociólogo dispõe dos resultados oficiais das eleições que lhe dão a repartição dos sufrágios por circunscrição. Essa repartição varia naturalmente segundo as circunscrições e ele pode procurar explicar essas variações a partir das diferenças entre as circunscrições que outras fontes lhe permitirão discernir. Assim, os resultados do recenseamento mais próximo fornecer-lhe-ão certas informações sobre a estrutura sócio-profissional das comunas. Poderá, então, confrontar estas informações com os resultados da consulta.

Tomemos um exemplo simples: imaginemos que eu queira relacionar o número de votos obtidos pela esquerda e a proporção dos operários. Disponho, para cada comuna, de dados que posso representar pelo quadro seguinte:

## Comuna X

|                             |                              |                               |                |
|-----------------------------|------------------------------|-------------------------------|----------------|
|                             | Número de operários          | Número de não-operários       |                |
|                             | de operários                 | na população                  |                |
| Número de votos de esquerda | $N_{OE}^-$ { 7.000<br>0      | $N_{OE}^-$ { 7.000<br>0       | $N_E = 7.000$  |
|                             | $N_{OE}^-$ { 3.000<br>10.000 | $N_{OE}^-$ { 20.000<br>13.000 | $N_E = 23.000$ |
|                             | $N_O = 10.000$               | $N_O = 20.000$                |                |

Ou dito de outra maneira, conheço os números  $N_E$  (número de votos de esquerda),  $N_E^-$  (número dos votos das outras categorias),  $N_O$  (número de operários),  $N_O^-$  (número de não-operários). Mas não conheço os números representados no interior do quadro pelos símbolos  $N_{OE}^-$  (número de operários que votaram na esquerda),  $N_{OE}^-$  (número de não-operários que votaram na esquerda) etc. é, igualmente, impossível deduzi-los do quadro. Na verdade, supomos  $N_O = 10.000$ ,  $N_O^- = 20.000$ ,  $N_E = 7.000$ ,  $N_E^- = 23.000$ . De um ponto de vista aritmético, os 7.000 votos de esquerda podem ser, teóricamente, tanto o feito dos operários ( $N_{OE}^- = 7.000$ ,  $N_{OE}^- = 0$ ) quanto dos não-operários ( $N_{OE}^- = 7.000$ ,  $N_{OE}^- = 0$ ). Naturalmente trata-se, aqui, de hipóteses extremamente pouco verossímeis. Mas ilustram as dificuldades lógicas com que se choca o sociólogo quando utiliza dados já "prontos".

A sondagem tem a vantagem de eliminar este gênero de dificuldades, pois que permite conhecer o voto emitido por um indivíduo e, ao mesmo tempo, a categoria sócio-profissional a que pertence. Pode-se, então, neste caso, avaliar sem dificuldade os números  $N_{OE}^-$ ,  $N_{OE}^-$ ,  $N_{OE}^-$ ,  $N_{OE}^-$ , que aparecem no interior do quadro.

A técnica de sondagem apresenta, sem dúvida, inconvenientes. Inicialmente, porque a população

observada é, por definição, restrita. Além disso, é possível que distorções se introduzam e uma fração não-negligenciável das respostas não sejam válidas. Mas esses inconvenientes são relativamente menores, pois as distorções que pode engendrar a técnica de sondagem podem, geralmente, ser detectadas.

Um outro inconveniente evidente do recurso aos dados já "prontos" é que se pode não encontrar nelas nenhuma informação sobre variáveis importantes para a análise de um fenômeno. Assim, posso emitir a hipótese de que os operários têm mais tendência a votar na esquerda quando suas esperanças de promoção são mais reduzidas. Naturalmente, os dados oficiais não me permitirão observar esta variável: "esperança de promoção". Em compensação, a pesquisa por sondagem poderá, se nela introduzo questões pertinentes, me dar informações a esse respeito.

Compreende-se o interesse metodológico das enquetes por sondagem. Permitem, de uma parte, obter elementos de uma população informações espartizadas e, por conseguinte, comparáveis de um indivíduo ou elemento a outro. De outra parte, permitem observar todas as variáveis introduzidas no nível das hipóteses e estabelecer diretamente as relações entre essas variáveis.

Há, bem entendido, circunstâncias em que a enquete por sondagem é impossível ou pouco recomendável. Seria absurdo recorrer a ela para analisar o funcionamento de uma sociedade restrita e impossível utilizá-la num estudo de sociologia histórica, por exemplo.

2. Tipos de enquetes. — As formas de enquetes diferentes das por sondagem colocam problemas metodológicos particulares que não podemos abordar no quadro deste volume. Trataremos, então, essencialmente, na seqüência deste capítulo das enquetes por sondagem, se bem que os problemas me-

todológicos levantados a seu propósito se possam aplicar em certos casos às outras formas de enquetes.

Pode-se, inicialmente, distinguir útilmente dois tipos de enquetes por sondagem. O primeiro tipo corresponde às sondagens atômicas; o segundo, às sondagens contextuais.

A sondagem atômica só permite, em princípio, construir variáveis individuais. As sondagens eleitorais são, em sua maioria, deste tipo. As informações que aí se encontram concernem, por exemplo, à escolha eleitoral do interrogado, à categoria sócio-profissional a que pertence, sua idade, seu sexo, etc. Por outro lado, aí se encontram informações que permitem construir variáveis mais complexas; por exemplo, o nível de interesse pela política, a atitude relativa às grandes questões políticas, ou, para retomar um exemplo já utilizado, o sentimento de sucesso pessoal, as chances de promoção percebidas, etc. É possível que todos estes fatos tenham uma influência sobre o comportamento eleitoral. Mas o importante é que todas estas variáveis são definidas ao nível do indivíduo. Caracterizam, pois, o indivíduo enquanto tal, mas dão uma informação limitada sobre o meio social a que pertence. Mais precisamente, este tipo de sondagem restringe o número das características do meio que é possível construir. Este inconveniente está longe de ser negligenciável, pois o que interessa em primeiro lugar ao sociólogo é, precisamente, determinar os fatores sociais do comportamento.

As sondagens contextuais são, por contraste, as que permitem construir variáveis que caracterizam não somente os indivíduos mas o meio a que pertencem. Assim, em uma enquete sobre as eleições no interior do sindicato americano dos tipógrafos, Lipset e seus colaboradores construíram um plano de observação de três níveis de amostragem. Antes de mais nada, estabeleceram uma amostra de seções locais e, no conjunto das seções locais, uma amostra

de seções de empresa. No interior das seções de empresa, não houve propriamente amostragem, pois foi decidido interrogar todos os indivíduos que pertencem às seções selecionadas.

Observemos de imediato que esse plano de observação particular não se impunha de forma alguma. Teria sido lícito proceder a uma amostragem direta, isto é, pegar a lista dos sindicalizados e escolher aleatoriamente nessa lista um subconjunto de indivíduos.

Vejamos, entretanto, em um exemplo simples, o que se teria perdido.

O interesse do plano de observação adotado era que permitiu construir variáveis características, não somente dos indivíduos mas, também, das seções. Assim, observando os votos emitidos em uma eleição em que se tratava de escolher entre dois candidatos, os autores puderam caracterizar as seções em função do grau de consenso que deixavam aparecer. Percebeu-se, então, que esta característica das seções estava fortemente ligada a certas variáveis individuais, como mostra o quadro seguinte. Nêle se vê, na verdade, que o grau de implicação dos indivíduos em relação aos problemas sindicais está ligado ao grau de consenso que caracteriza a seção à que pertencem (quadro I).

Este quadro pode ser interpretado de várias maneiras. Pode, na verdade, querer dizer ou que o consenso acarreta a implicação individual, ou que a implicação individual acarreta o consenso, ou que os dois efeitos estão simultaneamente presentes. Qualquer que ele seja, constata-se aqui que um comportamento individual é fortemente dependente de uma variável característica do meio.

\* S. M. LIPSET, M. THORP e J. S. COLEMAN, *Union democracy*, Glencoe, The Free Press, 1956.

QUADRO I  
RELAÇÃO ENTRE IMPLICAÇÃO SINDICAL E CONSENSO

| Porcentagem de indivíduos que manifestam uma implicação elevada em relação aos problemas sindicais ..... | Número de seções .....                               |  |
|--|--|--|
|  | Seções caracterizadas por um grande consenso elevado | Seções caracterizadas por um grande consenso baixo |
| 29%  | 7%   |  |
| (125)  | (28)   |  |

Naturalmente, este fenômeno não poderia ter sido demarcado se se tivesse utilizado uma sondagem de tipo atômico. Ter-se-iam, sem dúvida, encontrado variáveis individuais dando conta da implicação sindical. Mas se teria deixado escapar de suas determinantes essenciais: o clima político do meio.

As sondagens contextuais são chamadas a representar um papel fundamental na pesquisa sociológica. Na verdade, permitem analisar o comportamento individual colocando-o em uma "estrutura social", enquanto que as sondagens atômicas, como se lhes recriminou muitas vezes,\* consideram indivíduos destacados de seu contexto e colocados, por assim dizer, em um espaço social amorfo.

Naturalmente, o exemplo que demos mais acima é extremamente simplificado. Em uma sondagem contextual real, o meio social será caracterizado, não por uma variável, mas por um conjunto de variáveis de que se tratará de separar as relações recíprocas ao mesmo tempo que as relações que mantêm com as variáveis individuais.

Notemos, ainda, entre parênteses, que existe não uma mas várias formas de planos de observação contextuais. A variável contextual apresentada mais

\* HERBERT BRUMER, "Public opinion and public opinion polling", in Daniel Katz e col., *Public opinion and propaganda*, New York, Henry Holt, 1960.



acima (consenso político) é construída a partir dos votos individuais. Mas pode-se, também, utilizando por exemplo técnicas de tipo sociométrico, construir variáveis *relacionais* que permitam caracterizar a estrutura das relações entre os melhores de um grupo.

Esse tipo de variáveis é utilizado notadamente por J. Coleman em seus estudos de sociologia da educação.<sup>3</sup> Elas permitiram constatar que, independentemente do sucesso escolar e das características individuais habitualmente introduzidas nos estudos de mobilidade das causas saídas prematuras por parte das crianças provenientes notadamente das camadas desfavorecidas era sua ausência de integração no grupo que constituem os alunos de uma sala.

Em certos estudos, achou-se mesmo vantajoso adotar um plano de amostragem em "bola de neve". Nesta técnica, começa-se por definir uma amostra restrita comportando um pequeno número de pessoas. Somam-se em seguida as pessoas com as quais estas últimas declararam manter relações e continua-se, assim, a completar a amostra. Este plano pode ser aplicado quando se estudam meios muito vastos para que todos os indivíduos possam ser observados. Permite, então, determinar, com menor custo, as interdependências entre variáveis individuais e variáveis relacionais.

Se os planos de observação contextuais são relativamente pouco utilizados hoje, na França pelo menos, é porque representam um custo evidentemente mais elevado que as sondagens atômicas. Mas sua importância é tal para a análise sociológica que se pode prever que terão no futuro uma grande extensão.

<sup>3</sup> JAMES S. COLEMAN, *The adolescent society*, Glencoe, The Free Press, 1961.

Existe uma outra distinção importante entre as enquetes por sondagem: é a distinção entre sondagens *instantâneas* e sondagem por *painel*.

A sondagem instantânea é praticada em um momento dado e analisado enquanto tal. A sondagem por painel consiste em observar a mesma amostra em algumas retomadas. Apresenta um interesse fundamental na análise da mudança, pois permite seguir as mudanças de estado dos elementos de um sistema social de um momento a outro.

A técnica do painel foi principalmente aplicada no domínio da sociologia política em que permitiu analisar notadamente os mecanismos de cristalização das opiniões no decorrer dos períodos eleitorais. Mas ela é, igualmente, um elemento de importância capital para a análise dos fenômenos de mobilidade profissional e de mobilidade social.

Apresenta, entretanto, dificuldades que restringem seu uso.

Primeiro, é de aplicação custosa. Além disso, chama-se com o grave problema da "mortalidade" da amostra, problema tão mais grave quanto maior é o período abarcado pela observação.

Naturalmente, podem-se imaginar diferentes combinações entre as duas distinções apresentadas. Assim, pode-se imaginar sondagens por painel de tipo contextual, que permitiriam estudar a mudança individual em função da mudança da estrutura social ambiente.

3. A escolha da população-mãe. — Digamos finalmente uma palavra sobre um problema essencial na constituição do plano de observação: o da escolha da população-mãe da qual a amostra vai ser tirada.

O que importa, principalmente, compreender aqui é que o problema se apresenta de maneira inteiramente diferente na medida que uma enquete vise à *descrição* ou à *exploração*. Assim, se se deseja estimar a proporção dos médicos que utilizam, num momento dado, na França, tal medicamento novo ou a proporção dos franceses que, num momento dado, estão a favor da política econômica do gover-

no, a população-mãe que a amostra deverá representar será necessariamente definida como o conjunto dos médicos franceses no primeiro caso e como o conjunto dos franceses acima de uma certa idade no segundo.

Suponhamos agora que nos interessam os mecanismos de difusão social e nos perguntamos *por que* certos médicos adotaram o novo medicamento e outros não. Neste caso, a enquête visará responder a certas questões teóricas ou verificar certas hipóteses. Assim, poderemos supor que os médicos cujos contatos profissionais estão limitados a uma clientela privada estarão menos prontos a adotar uma novidade que os médicos dos hospitais, que são estimulados por mecanismos de influência interpessoal.

É evidente que, neste caso, seria inteiramente inútil tirar uma mostra representativa da população dos médicos franceses, pois se a relação que conto observar em virtude de minha hipótese é verificada numa amostra tirada dos meios médicos de uma cidade X, não há muita chance, aparentemente, para que não seja igualmente verificada numa cidade Y.

Vê-se, por êste exemplo, que a definição da população-mãe obedece a coações diferentes segundo a natureza lógica das questões a que se quer responder. Se se trata de verificar a existência de certas relações entre tal e tal variável, pode-se escolher uma população-mãe particular com a condição de ter a garantia de que êste contexto particular não corre o risco de afetar as relações em causa. No caso de questões de tipo descritivo, a população-mãe é, ao contrário, imposta.

### III. A construção das variáveis

Consideremos uma proposição sociológica qualquer. Por exemplo, uma das seguintes: o moral dos operários é pior em um sistema de organização autoritário; "o sucesso escolar depende estreitamente da aptidão em manejar a lingua de idéias próprias ao ensino";<sup>10</sup> a aceitação pelo imigrante da sociedade que o acolhe é mais fácil se êle pertence a uma família so-lidária; "o individualismo excessivo não tem somente por resultado favorecer a ação das causas suicidogê-nas, é, por si mesmo, uma causa dêste gênero".<sup>11</sup>

<sup>10</sup> PIERRE BOUROJEU e JEAN-CLAUDE PASSERON, *Les Héritiers*, Paris, Editions de Minuit, 1964.  
<sup>11</sup> DURKHEIM, *Le suicide*.

Estas proposições, tomadas de empréstimo a diversos autores, fazem aparecer, imediatamente, uma dificuldade. Para demonstrar, por exemplo, a proposição que liga a solidariedade familiar e a adaptação na sociedade que acolhe o imigrante, é preciso tirar uma amostra de imigrantes, classificá-los em relação aos dois critérios ou variáveis e demonstrar que há uma relação entre êstes critérios. Mas, como definir êstes critérios de maneira a atingir uma classificação razoável? Como, em outros termos, decidir de maneira não-equivoca que um emigrante está melhor adaptado que outro, que uma família é mais solidária que outra?

1. **Dos conceitos aos índices.** — Qualquer que seja o problema sociológico que se coloque ou a hipótese que se queira demonstrar, confrontar-nos sempre com o problema da *construção das variáveis*, isto é, da tradução dos conceitos e noções em operações de pesquisa definidas. A palavra "variável" tem, notemos de passagem, uma história ambígua. Originada da matemática e da física teórica, tomou nas ciências sociais um sentido cada vez mais amplo e pouco a pouco se estabeleceu o hábito de compreender no conceito de variável o resultado da partição de conjuntos de objetos segundo um ou vários critérios específicos: *sexo, nível de qualificação, idade* são exemplos de variáveis. As classificações que correspondem à primeira e à segunda são, respectivamente, nominal e ordenada; somente a terceira é quantitativa e do mesmo tipo que as variáveis utilizadas pela física. Poder-se-ia dizer para salientar as distinções que o sexo é um *atributo* de dois valores, a hierarquia profissional, uma *ordem*, a idade uma *variável no sentido estrito* da palavra. Mas, a maioria das vezes, fala-se de variável para designar um critério qualquer de classificação, quer êste conduza à definição de classes simplesmente distintas (ou *categorias*), de classes

ordenadas (ou posições), ou classes definidas por um valor quantitativo (ou medidas).

O problema da construção das variáveis é, pois, o da tradução dos conceitos em índices. Trata-se, em outros termos, de passar da definição abstrata ou da conotação intuitiva das noções sociológicas ("individualismo", "solidariedade familiar", etc.) a critérios que permitam definir uma classificação a partir dessas variáveis.

Eis como Lazarsfeld descreve as quatro etapas dessa passagem:<sup>12</sup>

"1º A *representação imaginada do conceito*. — Um problema clássico da sociologia industrial é a análise e a "medida" da noção de gestão. Mas que se entende exatamente por "gestão", "direção" e "administração"? O conceito pode ser considerado como agente de gestão? O conceito de gestão talvez tenha aparecido no dia em que se observou que duas fábricas, colocadas em condições idênticas, podem ser bem ou mal dirigidas. Este fator complexo, favorecendo o rendimento dos homens e a produtividade do equipamento, foi, então, identificado pelo nome de "gestão". A partir daí, os sociólogos das organizações se esforçaram por precisar essa noção e dar-lhe um conteúdo mais concreto.

A mesma evolução se manifestou em outros domínios. Hoje, o uso dos testes de inteligência tornou-se constante. Mas a noção de "inteligência" corresponde, originalmente, a uma impressão complexa e concreta de vivacidade ou de entorpecimento mental. E' muito freqüente uma impressão geral dessa ordem que desperta a curiosidade do pesquisador e o orienta em um caminho que conduz finalmente a um problema de medida.

"2º A *especificação do conceito*. — A fase seguinte consiste em analisar os "componentes" dessa primeira noção, que chamaremos ainda, segundo o caso, "aspectos" ou "dimensões". Pode-se deduzi-los analiticamente do conceito geral que os engloba, ou, empiricamente, da estrutura de suas intercorrelações. De qualquer forma, um conceito corresponde, geralmente, mais a um conjunto complexo de

fenômenos do que a um fenômeno simples e diretamente observável.

"Suponhamos que se deseje saber se o rendimento de uma equipe de operários é satisfatório. Só se tem, de início, uma noção bastante vaga do que é um rendimento satisfatório e perguntar-se-á, sem dúvida, o que implica tal expressão. Qual tipo de rendimento é necessário preferir: o de um operário que trabalha rápido e estraga muitas peças ou o de um operário lento mas cuidadoso em seu trabalho? Em certos casos, segundo a natureza da fabricação, pode-se admitir um rendimento medíocre associado a uma baixa taxa de perdas, parece, entretanto, pouco provável que, levado este raciocínio ao extremo, aceite-se eliminar completamente os riscos de erros pela adoção de uma cadência excessivamente baixa. Finalmente, somos levados a analisar a noção de rendimento e a determinar seus diferentes componentes: rapidez de trabalho, qualidade do produto, rentabilidade do equipamento. A teoria da medida dá a estes fatores o nome de "dimensões", cuja análise é muitas vezes um problema complexo, como se pode ver em estudo sobre uma fábrica de construção aeronáutica onde se puderam destacar dezenove componentes da noção de gestão. Eis alguns exemplos desses componentes: ausência de dissensões no grupo, boas comunicações hierárquicas, flexibilidade da autoridade, política racional da direção, importância relativa do efetivo dos quadros.

"3º A *escolha dos indicadores*. — A terceira etapa consiste em encontrar indicadores para as dimensões guardadas. Isto é fácil. A primeira dificuldade pode ser assim formulada: o que é exatamente um indicador? William James escrevia em *The meaning of truth*: "... Quando se diz que um homem é prudente, quer-se dizer com isso que ele adota um certo número de comportamentos característicos da prudência: que faz seguros, não aposta tudo no mesmo cavalo, não se lança sem refletir em uma empresa... O termo "prudente" é, assim, uma maneira prática de exprimir abstratamente um traço comum a seus atos habituais... Há no seu sistema psicofísico caracteres distintivos que o levam a agir prudentemente..."

Aqui, a tentativa de James vai de uma imagem a um conjunto de indicadores, diretamente sugeridos pela experiência da vida quotidiana. Tem-se hoje tendência a especificar a relação desses indicadores à qualidade fundamental: não se exige de um homem prudente que antes de jogar, reparta toda sua aposta com tanto cuidado, ou que se

<sup>12</sup> In *Le vocabulaire des sciences sociales* (cf. Bibliografia).

precaute contra todos os riscos que corre. Diz-se, sòmente, que é provável que executará certos atos específicos da prudência. Sabemos, também, que os indicadores utilizáveis variam muito segundo o meio social do indivíduo. Praticamente não se encontram ocasiões de apostar ou tratar uma apólice de seguros em um pensionato religioso, por exemplo. E, entretanto, possível elaborar uma medida da prudência que dê conta desse meio periclar.

Definida em termos de probabilidade e não de certeza a relação entre cada indicador e o conceito fundamental, é indispensável utilizar tanto quanto possível, um grande número de indicadores. O estudo dos testes de inteligência, por exemplo, permitiu decompor esta noção em várias dimensões: inteligência manual, verbal... Mas estas dimensões não podem ser medidas senão por um conjunto de indicadores.

"4.º A formação dos índices. — A quarta fase consiste em fazer a síntese dos dados elementares obtidos no decorrer das etapas precedentes. Tendo decomposto o rendimento de uma equipe de operários ou a inteligência de uma criança em seis dimensões, e escolhido dez indicadores para cada dimensão, trata-se, agora, de construir uma medida única a partir dessas informações elementares".

Para ilustrar este procedimento, tomemos um exemplo já utilizado: o da *The Authoritarian Personality*. Trata-se, lembramos, de demonstrar que o antisemitismo e, mais geralmente, o etnocentrismo, são componentes de uma síndrome de personalidade. Por isso os autores construíram uma variável-síndrome, à que deram o nome de *autoritarismo*. As dimensões desta noção são, segundo os autores, em número de nove. Foram deduzidas, por uma análise temática, das entrevistas, de que uma amostra está presente no primeiro parágrafo deste capítulo. A lista dessas dimensões é a seguinte:

1. Apêgo às convenções;
2. Submissão a um princípio de autoridade superior;
3. Agressividade autoritária (tendência a rejeitar as pessoas que não seguem as convenções);
4. Antitraceção (condenação de um desenvolvimento demasiado da subjetividade, da imaginação);
5. Tendência à superstição e aos estereótipos;
6. Insistência na oposição autoridade-obediência;
7. Espírito de destruição;
8. Pessi-

misimo (tendência a acreditar que catástrofes ameaçam o mundo) 9. Moralismo sexual.

Trata-se, em seguida, de construir indicadores que permitiam classificar os assuntos observados em cada uma de suas dimensões. Eis, a título de exemplo, algumas das proposições que serviram de "indicadores" para medir a "submissão a uma autoridade superior". Estas proposições foram livremente traduzidas. Os interrogados deviam, a propósito de cada uma delas, declarar seu acôrdo ou desacôrdo.

— É desprezível quem quer que não manifeste devoção, gratidão e respeito em relação a seus pais. — É indispensável para eficácia do ensino ou do trabalho que os professores ou superiores deem instruções precisas sobre as tarefas que propõem. — Todos deveriam crer em uma força sobrenatural superior. — Ciências como a medicina, a física ou a química fizeram progressos consideráveis, mas existem numerosas questões essenciais a que o espírito humano não pode responder. — A obediência e o respeito são as principais qualidades que é preciso ensinar a uma criança, etc.

O número de respostas aprovatórias ou desaprovatórias dadas permite, então, classificar os interrogados na dimensão considerada.

Anteriormente, um estudo de *validação* fôra efetuado: os autores tinham verificado que as correlações entre o resultado global e as respostas a cada um dos indicadores eram suficientemente elevadas para que estes pudessem efetivamente ser considerados como pertencentes a uma dimensão única.

Observemos que os princípios descritos pelo texto de Lazarfeld são os resultados, não de uma metodologia *a priori*, mas de uma reflexão sôbre as pesquisas efetuadas pelos sociólogos. Descrevem, de qualquer forma, os procedimentos para os quais tende, mais ou menos idealmente todo sociólogo envolvido numa pesquisa empírica.

A lógica descrita por esse texto está presente, ainda que sob uma forma ligeiramente diferente, não somente nos trabalhos recentes, mas em obras clássicas. Assim, quando Durkheim tenta demonstrar que o individualismo, ou como êle diz ainda, o *egoísmo* é uma das causas do suicídio, êle se choca com uma dificuldade de que o texto do *Suicide* mostra que está plenamente consciente. Essa dificuldade reside na passagem da definição abstrata de egoísmo à definição de critérios objetivos de egoísmo. Em resumo, trata-se de traduzir um conceito em variável. Para tanto, recorre a uma série de *indicadores*, de que examina separadamente a ligação estatística com as taxas de suicídio. Um desses indicadores é a oposição solteiro-casado. O solteiro, mais livre para fixar seu modo de vida em função de seus próprios desejos, está menos submetido às coações sociais que o espôso e, portanto, mais "egoísta". Da mesma forma, o pai de família é colocado mais diretamente sob o olhar da sociedade que o casado sem filhos. Essa distinção fornece, então, um outro fimido pela oposição "protestante-católico", pois, se os protestantes se suicidam mais, é, segundo Durkheim, porque sua religião os convida a procurar suas regras de conduta em si mesmos, enquanto que os católicos recebem-nas de uma autoridade moral concebida como exterior ao indivíduo.

Poder-se-ia, dessa forma, estabelecer uma lista relativamente longa dos indicadores de egoísmo utilizados por Durkheim. Uma lista análoga corresponderia à noção de *anomia*. A principal diferença entre os procedimentos de "medida" adotados por Durkheim e os que descreve Lazarfeld é que os indicadores utilizados no *suicídio* não são combinados entre si.

Mas é preciso notar que um obstáculo intransponível tornava essa combinação impossível. Na verdade, as estatísticas utilizadas por Durkheim permi-

tiam-lhe, por exemplo, determinar as taxas de suicídio para os solteiros e os casados, por um lado, para os protestantes e os católicos, por outro. Mas em nenhum caso elas lhe permitiam determinar as taxas de suicídio correspondentes às combinações desses critérios (protestantes solteiros, casados; católicos solteiros, casados). Mesmo se tivesse observado, a natureza da informação de que dispunha, vetava-lhe proceder a uma combinação dos indicadores.

Em outros casos, o ideal descrito pelo texto de Lazarfeld é ainda mais longínquo, mas, assim mesmo, presente. Consideremos, por exemplo, os trabalhos de Thomas e Znaniecki<sup>13</sup> e da escola de Chicago<sup>14</sup> sobre as causas da "desorganização social" e da "desmoralização individual" observadas entre os emigrados.

Nesses estudos, os elementos utilizados são de natureza qualitativa (histórias de vida, cartas, descrição "etnológica" das relações sociais, dos conflitos, etc.). Em nenhuma parte, as *relações* entre as variáveis introduzidas ("desorganização social", "desmoralização", etc.) dão lugar a uma análise formal. Entretanto, seria fácil mostrar que estes autores utilizam bem, ainda que de maneira implícita, a linguagem dos indicadores, das variáveis, e das relações entre variáveis. Assim, o aparecimento frequente de condutas erráticas entre os imigrantes (que se vê deixarem sua família para voltar ao fim de algumas semanas, fugirem de nôvo, pedirem o divórcio, suspenderem abruptamente a ação encetada, etc.) é interpretada como um indicador de "desmoralização", em geral associado a outros indicadores, como a ausência de planos profissionais ou o desaparecimento da autoridade do pai sobre as crianças. Assim, mesmo no nível desses estudos de ordem qualitativa, vêem-se aparecer em filigrana os indi-

<sup>13</sup> *The Polish peasant...* op. cit.  
<sup>14</sup> Ver Madge, *The origins...* (cf. Bibliografia).

cadores sob os conceitos e as relações entre variáveis sob as descrições monográficas.

Veremos mais abaixo que se procurou dar uma formulação matemática ao problema da construção das variáveis a partir dos indicadores. E' por isso que seria bom sublinhar que êsse problema não é a invenção de uma metodologia *a priori* mas se enraíza na pesquisa sociológica mais concreta e, às vezes, a mais qualitativa.

2. A intermutabilidade dos índices. — Evidentemente não é possível definir, em sociologia, medidas que têm o mesmo grau de validade que as utilizadas nas ciências da natureza. Os autores de *The Authoritarian Personality* mostraram que a submissão ao princípio de autoridade se relacionava muito com o etnocentrismo. Durkheim aparentemente estabeleceu que o egoísmo era uma causa do suicídio. Mas êstes resultados não se tornaram duvidosos pelo fato de que a escolha dos indicadores é sempre, de certa maneira, arbitrária? Na verdade, poder-se-ia imaginar mil outros critérios de submissão ao princípio de autoridade ou de egoísmo além dos que foram efetivamente utilizados.

Digamos de imediato que não existe nenhuma resposta satisfatória a esta dificuldade no plano da lógica. A única maneira de se assegurar que os indicadores escolhidos não dão lugar a resultados artificiosos ou fortuitos é, em caso de dúvida, escolher outros indicadores e verificar se dão os mesmos resultados. E' o que faz Durkheim quando analisa as relações entre as taxas de suicídio, por um lado, e uma multiplicidade de indicadores de "egoísmo" ou de "anomia", por outro.

Eis, entretanto, uma experiência metodológica, igualmente devida a Lazarfeld, de onde se deduz que não se deve exagerar os inconvenientes dessa situação.

Tratava-se de analisar a relação entre a promoção ao grau de professor titular, por um lado, e a idade e a eminência científica, por outro.<sup>15</sup> Mas como medir o grau de eminência? Escolheram-se para tanto duas baterias de indicadores diferentes. A primeira levava em consideração os cargos científicos exercidos e os títulos universitários. A segunda utilizava indicadores de produtividade (número de livros, de artigos publicados, etc.). As duas baterias de indicadores conduziram, como se podia esperar, a classificações muito diferentes dos professores, já que 36% dos indivíduos observados foram classificados de maneira distinta pelos dois índices. Entretanto, quando se analisou a estrutura das relações entre a variável "eminência" medida por cada um dos dois índices, por um lado, e as variáveis "idade" e "promoção" por outro, observou-se que essa estrutura era idêntica, qualquer que seja o índice de "eminência" utilizado.

Eis, na verdade (quadro II), a percentagem de professores titulares para cada categoria de idade e cada classe de eminência (a eminência sendo definida na primeira metade do quadro pelos cargos científicos e os títulos universitários e, na segunda, pela produtividade).

QUADRO II

ESTRUTURA DAS RELAÇÕES CORRESPONDENTE A DUAS DEFINIÇÕES DA VARIÁVEL "EMINÊNCIA"

|                                     | Menos de 40 anos | De 41 a 50 anos | Mais de 50 anos |
|-------------------------------------|------------------|-----------------|-----------------|
| <i>Eminência (cargos, títulos):</i> |                  |                 |                 |
| Alta                                | 18 % (312)       | 65 % (308)      | 88 % (368)      |
| Média                               | 6 — (298)        | 28 — (149)      | 73 — (148)      |
| Baixa                               | 2 — (488)        | 22 — (150)      | 44 — (132)      |
| <i>Eminência (produtividade):</i>   |                  |                 |                 |
| Alta                                | 15 % (324)       | 63 % (358)      | 87 % (421)      |
| Média                               | 7 — (349)        | 39 — (131)      | 65 — (122)      |
| Baixa                               | 2 — (439)        | 23 — (126)      | 45 — (108)      |

<sup>15</sup> Le vocabulaire des sciences sociales (cf. Bibliografia).

As percentagens variam sensivelmente de um quadro a outro. Entretanto, a *estrutura* das relações entre variáveis, que só importa do ponto de vista sociológico, é a mesma no quadro de cima e no de baixo, como o leitor poderá verificar traduzindo tais resultados de maneira gráfica.

Este exemplo mostra que, apesar das classificações obtidas por índices diferentes poderem ser muito diferentes, a estrutura das relações entre uma variável medida por índices e outras variáveis pode ser a mesma. Aliás, isso se explica facilmente, pois a instância que decide da promoção de um professor se encontra exatamente na mesma situação que o sociólogo que procura medir sua eminência: sua avaliação da eminência é igualmente efetuada com a ajuda de indicadores mais ou menos explícitos. Da mesma forma, o "egoísmo" pode assumir diferentes formas. Na medida em que essas formas traduzem uma realidade social única, pode-se prever que manifestem as mesmas relações com variáveis como a propensão ao suicídio.

#### IV. A análise das relações entre variáveis

As variáveis uma vez construídas, trata-se de analisar suas relações. Diversos estudos mostram, por exemplo, que o suicídio é mais freqüente entre os divorciados, o absentismo profissional mais freqüente entre as mulheres, o nível de aspiração mais baixo entre as crianças educadas em uma família de estrutura autoritária, as oportunidades de entrar no ginásio menores para as crianças oriundas de famílias pertencentes às classes desfavorecidas. Tais correlações são muitas vezes o ponto de partida da análise sociológica. Assim, todo *Le Suicide* de Durkheim é um esforço para explicar certas relações primitivas deste tipo: relação entre taxa de divórcio e taxa de suicídio, entre duração do dia e suicídio, etc.

Os estudos concernentes à mobilidade social partem sempre de relações entre comportamentos de mobilidade e categorias sócio-profissionais. E' o caso, por exemplo, dos trabalhos de A. Girard sobre a admissão no ginásio ou de P. Bourdieu e J.-C. Passeron sobre o acesso ao ensino superior.<sup>16</sup>

Na fase pré-durkheimiana da sociologia empírica, a tendência era tomar relações como aquelas, como certas, e interpretá-las diretamente. Esta prática repousava sobre o postulado errôneo segundo o qual uma correlação estatística entre duas variáveis seria sempre o sinal de uma relação de causalidade entre essas variáveis.

Assim, os positivistas italianos como Lombroso ou Ferri, interpretavam a relação que as estatísticas permitiam estabelecer entre taxas de suicídio e temperatura (as taxas de suicídio crescem, na verdade, com a temperatura) como uma relação de causalidade: o calor, diziam eles, provoca um estado de excitação física que aumenta a propensão ao suicídio. A desigualdade das oportunidades de sucesso escolar ou social é, da mesma forma, interpretada por autores como Galton e Pearson como produto de fatores hereditários.<sup>17</sup>

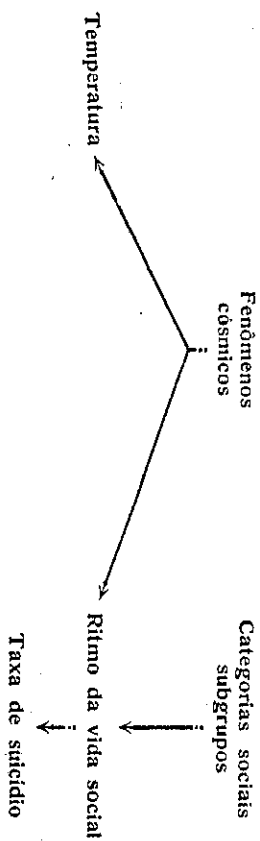
1. Explicação sociológica e sistemas de relações causais. — Durkheim foi o primeiro a ter compreendido que uma relação estatística só pode ser interpretada como relação causal com grandes precauções.

Tomemos o exemplo de sua análise da relação entre suicídio e temperatura. Os dias mais quentes, observa, são também os dias mais longos. Mas isso não indica que o suicídio é mais freqüente quando a vida social é mais longa? Efetivamente, as estatísticas mostram que as taxas de suicídio variam não somente com as estações, mas, também,

<sup>16</sup> Ver *Population*, 1963 e 1964 e *Les Héritiers*, op. cit.  
<sup>17</sup> ALAIN GIRARD, *La réussite sociale*, Paris, Presses Universitaires de France, coleção "Que-sais-je?"

com o ritmo da atividade que caracteriza os dias da semana bem como as horas da jornada de trabalho. Além disso, quando o tempo social varia com as categorias ou subgrupos de uma sociedade, as taxas de suicídio seguem essas diferenças de ritmo. Assim, a taxa de suicídio feminino se eleva durante o fim-de-semana, enquanto o suicídio masculino decresce nesse período; isto se origina de que — pelo menos na época em que escreveu Durkheim — o fim-de-semana corresponde ao nível máximo das atividades sociais da mulher e ao nível mínimo das atividades sociais do homem. Da mesma forma, o tempo social urbano era, nessa época, muito mais regularizado que o tempo social rural, que dependia mais estritamente do tempo astronômico. Este fato explica, segundo Durkheim, que a taxa de suicídio seja, em vários países observados, regularmente máxima no verão, enquanto nos centros urbanos desses mesmos países aparece ora na primavera, ora no verão. Além disso, enquanto que no conjunto dos países as taxas de suicídio crescem em 50%, aproximadamente, da estação em que o suicídio é menos frequente (o inverno) à que é mais frequente, não cresce senão 25%, no máximo, nas grandes cidades desses mesmos países. Este resultado traduz, novamente, a regularização do tempo social nas cidades e mostra que as variações da taxa de suicídio com a temperatura e com a extensão do dia devem ser explicadas pelo efeito do ritmo da vida social sobre o suicídio.

Esse tipo de análise, que se encontra em *Le Suicide*, pode ser, do ponto de vista lógico, considerado como verdadeira revolução. Mostra que uma relação estatística entre duas variáveis geralmente só pode ser interpretada se a inserimos em um modelo causal. Pode-se traduzir graficamente a análise durkheimiana da relação entre suicídio e temperatura (fig. 1).



Vê-se por este esquema que a relação entre temperatura e taxa de suicídio é explicada pela dupla influência dos fenômenos cósmicos sobre a temperatura e sobre o ritmo da vida social. Mas a explicação demonstra, ao mesmo tempo, que a temperatura não exerce nenhuma influência sobre a taxa de suicídio.

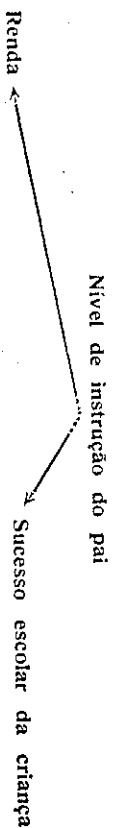
Das análises durkheimianas se tira a idéia de que a explicação de uma relação estatística consiste em introduzir variáveis suplementares, de maneira a liberar o modelo causal no qual se insere.

É um procedimento deste tipo que segue, por exemplo, A. Girard para explicar a relação observada entre renda dos pais e sucesso escolar.<sup>18</sup> Essa relação é positiva. O sucesso é, em outros termos, tão maior quanto a renda é mais elevada. Mas qual é a significação desse fato? É preciso interpretá-lo diretamente e admitir, por exemplo, que uma maior segurança econômica leva os pais a projetar estudos mais longos e a estimular mais o zelo escolar da criança? É uma interpretação possível. É corrigida, entretanto, pelo fato de que, quando se consideram crianças oriundas de famílias que apresentaram um nível cultural semelhante, a relação entre renda e sucesso escolar desaparece. Esse resultado indica que a relação entre renda e sucesso deve-se, de fato,

<sup>18</sup> *Population*, 1963 e 1964; *La réussite sociale*, op. cit.



a que um nível de instrução mais elevado correponde, em média, a um nível de renda mais elevado. Mas é, de fato, o nível cultural da família que é responsável pelo sucesso escolar da criança. Neste caso, a relação primitiva se encontra, então, explicada por um modelo que se pode representar pelo esquema da figura 2.



Este modelo mostra que a relação estatística entre renda e sucesso escolar não deve ser interpretada diretamente, mas é uma consequência do sistema de relações que caracteriza as três variáveis consideradas.

Assim como a noção de variável está presente em todos os estudos sociológicos, tanto nos estudos monográficos e qualitativos quanto nos estudos quantitativos, da mesma forma o postulado durkheimiano segundo o qual a análise sociológica consiste em determinar as estruturas causais que explicam as relações entre variáveis é comum à maioria das pesquisas sociológicas.

Mas é somente no caso em que se dispõe de medidas que correspondam a essas variáveis que se pode determinar com precisão o sistema de relações causais explicativo de um fenômeno social. Esse fato diz, sózinho, do desenvolvimento considerável das enquetes de tipo quantitativo.

**2. Os problemas lógicos da análise das relações entre variáveis: a análise multivariada.** — A análise das relações entre variáveis coloca problemas lógicos delicados, como se verá no capítulo seguinte. Examinaremos, por enquanto, com Lazars-

feld, as diversas situações que podem logicamente se apresentar quando se introduz uma variável suplementar em uma relação primitivamente observada entre duas variáveis.<sup>19</sup>

O exemplo escolhido por Lazarsfeld é muito simples: trata-se das relações entre a idade e o interesse por três tipos de transmissões radiofônicas. Essas relações estão presentes no quadro III.

**QUADRO III**

**RELAÇÕES ENTRE A IDADE E O INTERESSE POR TRÊS TIPOS DE TRANSMISSÕES RADIOFONICAS**

|                              | Jovens        |         | Velhos        |         |
|------------------------------|---------------|---------|---------------|---------|
|                              | % de ouvintes |         | % de ouvintes |         |
| Programas religiosos .....   | 17            | 26      | 45            | 29      |
| Tribunas políticas .....     | 34            | 45      | 29            | 29      |
| Programas de música clássica | 30            | 29      | 29            | 29      |
| (Número de casos) .....      | (1.000)       | (1.300) | (1.300)       | (1.300) |

Portanto, as pessoas mais idosas buscam mais os dois primeiros tipos de emissão. Em compensação as porcentagens são aproximadamente idênticas no que concerne ao terceiro tipo de programas.

Pode-se tentar interpretar diretamente essas relações. Uma teoria possível a propósito da primeira dessas relações seria de ver aí o sinal de uma ligação menor dos jovens aos fatores religiosos. A ausência de relação observada no terceiro caso poderia, por seu lado, ser interpretada pelo fato de que o interesse pela música é uma estrita questão de gosto.

Veremos que a introdução de uma variável suplementar, que Lazarsfeld chama de variável-teste, pode contribuir para esclarecer a interpretação. A variável-teste utilizada é o nível de instrução. Eis os resultados que se obtêm quando se analisam as

<sup>19</sup> "L'interprétation des relations statistiques comme procédure de recherches", in *L'analyse empirique de la causalité* (cf. Bibliografia).

relações precedentes repartindo a população observada em dois grupos caracterizados, respectivamente, por um nível de instrução "superior" e "inferior". Examinemos inicialmente o caso das emissões religiosas (quadro IV).

#### QUADRO IV

##### INTERESSE PELAS EMISSÕES RELIGIOSAS EM FUNÇÃO DA IDADE E DO NÍVEL DE INSTRUÇÃO

| Jovens<br>17%               |        | Velhos<br>26%               |        |
|-----------------------------|--------|-----------------------------|--------|
| Nível de instrução superior |        | Nível de instrução inferior |        |
| Jovens                      | Velhos | Jovens                      | Velhos |
| 9%                          | 11%    | 29%                         | 32%    |

Neste caso, vê-se que, homogeneizando os grupos pela relação ao nível de instrução, a relação primitivamente observada se atenua consideravelmente, a ponto de se tornar praticamente nula. Em outros termos, a relação entre a idade e o interesse por esse tipo de programas é devida a que, em uma sociedade submetida a um processo de democratização do ensino, os jovens têm, em média, um nível de instrução mais elevado. Como o interesse pelas emissões religiosas depende do nível de instrução (como se pode perceber comparando as colunas 1 e 3 e as colunas 2 e 4 do quadro IV), daí resulta uma relação entre a idade e o interesse pelas emissões religiosas. A relação desaparece quando se homogeneizam os grupos pela relação ao nível de instrução: a influência desta última variável é a única real. A análise pode ser resumida pela estrutura causal representada na figura 3.

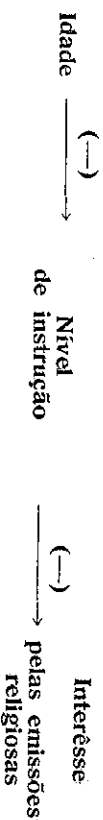


Fig. 3. Estrutura causal correspondente ao quadro IV.

O sinal (—) colocado acima das setas indica que a relação é negativa: se a idade é mais avançada, o nível de instrução é, em média, mais fraco; se o nível de instrução é mais elevado, o interesse pelas emissões religiosas é, em média, mais fraco.

Vê-se, em particular, que a relação não pode ser interpretada como tradutora de uma menor ligação dos jovens aos valores religiosos, pois que essa ligação é a mesma, qualquer que seja a idade, se o nível de instrução é o mesmo.

Observemos agora os resultados concernentes ao interesse pelas emissões políticas (quadro V).

#### QUADRO V

##### INTERESSE PELAS EMISSÕES POLÍTICAS EM FUNÇÃO DA IDADE E DA INSTRUÇÃO

| Jovens<br>34%               |        | Velhos<br>45%               |        |
|-----------------------------|--------|-----------------------------|--------|
| Nível de instrução superior |        | Nível de instrução inferior |        |
| Jovens                      | Velhos | Jovens                      | Velhos |
| 40%                         | 55%    | 25%                         | 40%    |

A situação é inteiramente diferente no caso. Na verdade, a relação entre a idade e o interesse persiste quando se homogeneizam os grupos pela relação à instrução. As duas variáveis, idade e instrução, têm, pois, um efeito sobre o interesse pelas emissões políticas. A estrutura causal correspondente está representada na figura 4.

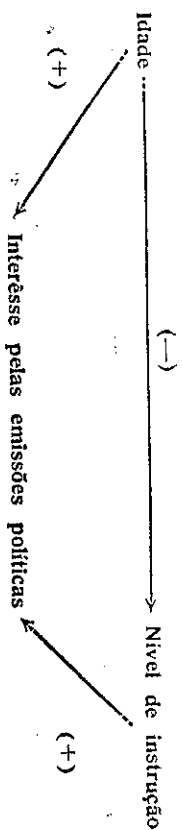


Fig. 4. Estrutura causal correspondente ao quadro V.

A introdução da variável-teste permite, ainda aqui, melhorar a interpretação. A instrução tem por efeito aumentar o interesse pela política, talvez porque induza no indivíduo o sentimento de que ele é capaz de compreender os acontecimentos políticos, que ele tem o direito de julgá-los, que pode agir sobre eles; Mas, talvez, também, o mais débil interesse das pessoas menos instruídas é devido a que ocupam, em média, posições sociais mais baixas, que experimentam um sentimento de exclusão e desenvolvem uma atitude de retraimento. Para decidir entre estas duas interpretações, seria necessário introduzir novas variáveis-testes. Quanto à idade, há um efeito independente do da instrução. Isso indica, talvez, que o tipo de inserção social do adulto acarreta nêle o sentimento de ser atingido diretamente pelos acontecimentos políticos.

No caso das emissões de música clássica, a estrutura das relações é ainda diferente, como o mostra o quadro VI.

#### QUADRO VI INTERESSE PELA MÚSICA CLÁSSICA EM FUNÇÃO DA IDADE E DA INSTRUÇÃO

| Jovens<br>30%               |        | Velhos<br>29%               |        |
|-----------------------------|--------|-----------------------------|--------|
| Nível de instrução superior |        | Nível de instrução inferior |        |
| Jovens                      | Velhos | Jovens                      | Velhos |
| 32%                         | 52%    | 28%                         | 19%    |

Neste caso, a relação primitivamente nula entre a idade e o interesse pelas emissões de música clássica torna-se *positiva* no grupo de instrução superior e *negativa* no grupo de instrução inferior. Pode-se imaginar que essa estrutura trai o fato de que nas pessoas de nível de instrução inferior, o momento em que o contato com a *cultura* é mais direto é o da escolaridade, em seguida passam a pertencer a meios sócio-profissionais em que a cultura clássica quase nunca é percebida como um valor central. Correlativamente, o fato de que a relação entre idade e interesse pela música clássica seja positiva no caso do grupo de instrução superior pode indicar que a música clássica é um valor cultural mais importante para as pessoas mais idosas, enquanto sofre mais concorrência de outras formas de cultura (cinema, música moderna, "jazz") no público jovem. Qualquer que seja, vê-se que a introdução da variável-teste evidencia fenômenos relativamente complexos, que era impossível destacar no nível da relação primitiva. A nulidade dessa relação é, vê-se, o produto de efeitos antitéticos que se compensam. A estrutura causal correspondente pode ser descrita pelos esquemas da figura 5.

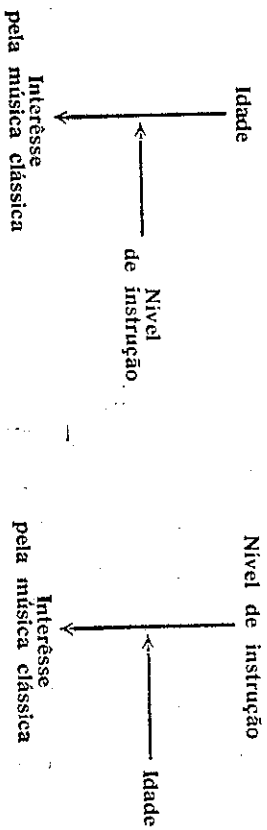


Fig. 5. Estruturas causais correspondentes ao quadro VI.

Esses esquemas sugerem que, neste caso, a influência da idade sobre o interesse pela música clássica depende do nível de instrução, e que, correlativamente, o efeito do nível de instrução sobre o interesse pela música clássica depende da idade. Vê-se que não podemos contentar-nos aqui em enunciar os efeitos das variáveis umas sobre as outras. É preciso, também, considerar efeitos de uma natureza mais complexa, que os estatísticos chamam efeitos de *interação*. No caso presente, há interação entre a idade e o nível de instrução, pois o efeito de cada uma destas variáveis sobre o interesse pela música depende da outra variável. Em outros termos, não há praticamente sentido em falar neste caso de efeito da idade sobre o interesse pela música clássica, pois este efeito difere completamente (mudança de sentido) caso se considerem as de instrução superior ou as pessoas de instrução inferior. Da mesma forma, não se pode falar de um efeito de instrução, pois ele varia com a idade. É precisamente esta impossibilidade de separar os efeitos das duas variáveis que indica a expressão "efeito de interação". É fácil demonstrar que os três tipos de estruturas ilustradas pelos exemplos precedentes esgotam as situações que se pode logicamente encontrar quando se analisa uma relação primitiva entre duas variáveis à luz de uma terceira variável. Vê-se, além disso, que a análise pode, neste caso, proceder de

maneira intuitiva. Em outros termos, é possível, quando se limita a análise a três variáveis, associar de maneira intuitiva uma estrutura causal interpretativa aos dados numéricos.

Mas a lógica mesma da análise (chamada análise *multivariada*) que acabamos de expor convida a generalizá-la. Vimos que a interpretação era consideravelmente enriquecida quando se introduzia uma terceira variável na análise de uma relação primitiva entre duas variáveis.

Mas, ao mesmo tempo, as estruturas com três variáveis sugerem novas variáveis intermediárias. Em plano mais geral, é natural estender a lógica da análise multivariada a quatro, cinco, e mesmo a um número de variáveis mais elevado ainda.

Tomemos um problema concreto já evocado. Imaginemos que, um estudo sobre a escolha profissional, se tenha submetido um questionário a uma população de alunos de ensino secundário. Certos alunos já têm idéia precisa de seu futuro profissional, outros não. Se o sociólogo se propõe a explicar este fenômeno, várias hipóteses vão-lhe parecer plausíveis. Talvez o incitamento para escolher seja maior nas famílias desfavorecidas. Talvez a escolha tenha mais chances de ser postergada nas famílias que mantêm a criança em um estado de proteção mais firme. Talvez o fato de pertencer a certas seções mais orientadas para a cultura geral tenha por consequência reduzir a escolha. Talvez a escolha tenha mais tempo de ser definida se o sucesso escolar é mais marcado em uma matéria que em outra, etc.

Poder-se-ia, assim, alongar a lista das hipóteses. O exemplo mostra que o sociólogo que analisa um fenômeno qualquer é normalmente levado a enumerar um conjunto de determinantes. Além disso, essas determinantes estão geralmente ligadas entre si. É bem conhecido, por exemplo, que a composição social das classes do "hiceu" varia com as seções: é menos provável fazer latim quando se per-